

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:02:33 ; Search time 190.526 Seconds
(without alignments)
118.199 Million cell updates/sec

Title: US-10-008-355-25
Perfect score: 54
Sequence: 1 TGCNSGSEPF 10

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/qgn2.1/USPQ/US10008355/runat_17122002_112336_14573/app.query.fasta_1.398
-DB=N.Geneseq_101002 -QFMT=fastap -SUFIT=p2n.rng -MINMATCH=0.1 -DOOPCL=0
-LOOPEXT=0 -UNITS=bits -STAR=1 -END=1 -MATRIX=Blomsum62 -TRANS=human40.cdt
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355 @CGN 1.1.0 @runat_17122002_112336_14573 -NCP=6 -ICPU=3
-NO_XIPXY -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WATP=LONGLOC -DEV=TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: N.Geneseq_101002.*
- 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	54	100.0	1019	AAV75063	Staphylococcus aur
2	54	100.0	1558	AAO27988	Protease from S. A
3	54	100.0	1586	AAO27987	Protease from S. A
4	54	100.0	2139	AAI43635	porphyromonas ging
5	49	90.7	849	AAH52485	S. epidermidis ope
6	49	90.7	936	ABN91429	Staphylococcus epi
7	49	90.7	3189	AAH54330	S. epidermidis gen
8	44	81.5	402	AAH17622	Human breast cance
9	44	81.5	423	AAI00017	Human reproductive
10	44	81.5	758	AAH20213	Enterococcus faeca
11	44	81.5	758	ABN98198	E faecalis EF110 g
12	44	81.5	888	AAH20212	Enterococcus faeca
13	44	81.5	888	ABN98197	E faecalis EF110 g
14	44	81.5	2421	AAV80633	Kidney injury asso
15	44	81.5	4951	AAH13198	Enterococcus faeca
16	44	81.5	17391	ABL30036	Drosophila melanog
17	44	79.6	534	ABK73036	Bacillus licheniflo
18	43	79.6	948	AAO29999	Bacillus licheniflo
19	43	79.6	1448	AAO24382	Protease Blase. B
20	42	77.8	152	AAH79971	Nucleotide sequenc
21	42	77.8	473	ABV56742	Human prostate exp
22	42	77.8	1546	ABO54166	Human ovary anti
23	42	77.8	319608	AAH51601	Human chromosome 1
24	42	77.8	319608	AAH09301	Human chromosome 1
25	41	75.9	2130	ABO91212	M. capsulatus gene
26	41	75.9	8922	AAK87069	Human immune/haema
27	40	74.1	370	ABN25944	Human OREX polynuc
28	40	74.1	450	ABL78727	Human ovarian can
29	40	74.1	942	AAO03001	Bacillus licheniflo
30	40	74.1	954	AAO03003	Bacillus licheniflo
31	40	74.1	1030	AAH99612	Human protein enco
32	40	74.1	1127	AAH07536	Fusarium venenatum
33	40	74.1	1873	AAH19209	DNA encoding human
34	40	74.1	7441	AAH29979	Human lung antigen
35	40	74.1	44861	AAH20000	DNA encoding pyrid
36	39	72.2	240	AAH89059	Human prostate can
37	39	72.2	584	ABA31488	Probe #9954 for ge
38	39	72.2	584	AAK12811	Human brain expres
39	39	72.2	741	AAH14488	Aspergillus oryzae
40	39	72.2	1173	AAK32891	DNA encoding C. al
41	39	72.2	1261	AAH04382	Human secreted pro
42	39	72.2	1303	AAH26291	Pseudomonas sp Typ
43	39	72.2	1775	ABT03070	Human breast speci
44	39	72.2	2012	AAH59942	Human secreted pro
45	39	72.2	2012	AAH33087	Human colon cancer

ALIGNMENTS

RESULT 1
AAV75063
ID AAV75063 standard. DNA: 1019 BP.
AAV75063;
16-MAR-1999 (first entry)
uuhw
fuuw

DE Staphylococcus aureus contig SEQ ID #752.
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX Staphylococcus aureus.
OS
XX
FH Key Location/Qualifiers

FT misc-feature 901..960
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
PN EP786519-A2.
XX 30-JUL-1997.
XX
XX 07-JAN-1997; 97EP-0100117.
XX
XX 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX
XX WPI: 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
XX stored on computer readable medium and used in the production of
XX anti-S.aureus vaccines
XX
XX Claim 1: Page 1646-1647; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX of the invention. The DNA sequences are recorded on a computer readable
XX medium, preferably selected from a floppy or hard disk, random access
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX the S.aureus DNA sequences allows putative functions to be assigned so
XX that protein-encoding or regulatory regions of commercial, therapeutic or
XX industrial importance can be obtained. Specifically, sequences which are
XX likely to encode antigens have been identified and these polypeptides can
XX be used in a vaccine composition against S.aureus infection. The
XX polypeptides can also be used in a kit for the immunodetection of
XX S.aureus in a sample. S.aureus is implicated in numerous human diseases,
XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX skin and surgical wound infections, scalded skin syndrome, toxic shock
XX syndrome, etc. Organisms transformed with the DNA sequences can be used
XX for recombinant production of the polypeptides. The new DNA sequences
XX (and their fragments) are useful as primers or probes for isolating
XX homologues of any of the S.aureus DNA sequences contained on the
XX computer readable medium.
SQ Sequence 1019 BP; 364 A; 194 C; 159 G; 240 T; 62 other:

Alignment Scores:
Pred. No.: 1.64 Length: 1019
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 18 Gaps: 0

US-10-008-355-25 (1-10) x AAV75063 (1-1019)
QY 1 ThrglyGlyAsnSerGlySerProValPhe 10
DB 642 ACTGGTGGTAATTCAGGTCACCTGTATT 671

RESULT 2
AAQ27988
ID AAQ27988 standard; DNA; 1558 BP.
XX
XX AAQ27988;
XX
XX 11-FEB-1993 (first entry)
XX
XX Protease from S. Aureus.
XX
XX

KW Protease; PCR; amplify; staphylococcus; ss.
XX
XX Staphylococcus aureus.
OS
XX
XX Key Location/Qualifiers
FH CDS 354..1364
FT /*tag= a
FT 354..557
FT sig_peptide /*tag= b
FT 558..1361
FT mat_peptide /*tag= c
XX
XX JP04211370-A.
XX
XX
XX 03-AUG-1992.
XX
XX
XX 19-FEB-1991; 91JP-0024633.
XX
XX 20-FEB-1990; 90JP-0040398.
XX
XX (SHIO) SHIONOGI & CO LTD.
XX
XX WPI: 1992-304938/37.
XX
XX P-PSDB; AAR29644.
XX
XX Novel protease prep. using Bacillus or Saccharomyces host -
XX capable of cleaving peptide bond at carboxyl terminus of glutamic
XX acid residues in polypeptide(s)
XX
XX Disclosure: Page 15-16; 25pp; Japanese.
XX
XX The sequences given in AAQ27987-88 encode proteases which were isolated
XX from Staphylococcus aureus strains. The DNA sequences were isolated
XX by PCR using the primer sequences given in AAQ27960-86. The protease
XX specifically cleaves the peptide bond at the C-terminus of the
XX glutamic acid residue in polypeptide.
XX
SQ Sequence 1558 BP; 579 A; 282 C; 221 G; 476 T; 0 other:

Alignment Scores:
Pred. No.: 2.63 Length: 1558
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-10-008-355-25 (1-10) x AAQ27988 (1-1558)
QY 1 ThrglyGlyAsnSerGlySerProValPhe 10
DB 1050 ACTGGTGGTAATTCAGGTCACCTGTATT 1079

RESULT 3
AAQ27987
ID AAQ27987 standard; DNA; 1586 BP.
XX
XX AAQ27987;
XX
XX 11-FEB-1993 (first entry)
XX
XX Protease from S. Aureus ATCC12600.
XX
XX Protease; PCR; amplify; Staphylococcus; ss.
XX
XX Staphylococcus aureus.
XX
XX
XX Key Location/Qualifiers
FH CDS 352..1425
FT /*tag= a
FT 352..555
FT sig_peptide /*tag= b
FT 556..1425
FT mat_peptide

```
FT      /*tag= C
XX
XX      JP04211370-A.
XX
XX      03-AUG-1992.
XX
XX      19-FEB-1991; 91JP-0024633.
XX
XX      20-FEB-1990; 90JP-0040398.
XX
XX      (SHIO ) SHIONOGI & CO LTD.
XX
XX      WPI; 1992-304938/37.
XX      P-PSDB; AAR26842.
XX
XX      Novel protease prep'd. using Bacillus or Saccharomyces host -
XX      capable of cleaving peptide bond at carboxyl terminus of glutamic
XX      acid residues in polypeptide(s)
XX
XX      Disclosure; Page 13-15; 25pp; Japanese.
XX
XX      The sequences given in AAQ27987-88 encode proteases which were isolated
XX      from Staphylococcus aureus strains. The DNA sequences were isolated
XX      by PCR using the primer sequences given in AAQ27960-86. The protease
XX      specifically cleaves the peptide bond at the C-terminus of the
XX      glutamic acid residue in polypeptide.
XX
SQ      Sequence 1586 BP; 590 A; 302 C; 219 G; 475 T; 0 other;

Alignment Scores:
Pred. No.:      2.68      Length:      1586
Score:          54.00      Matches:      10
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:            13      Gaps:      0

US-10-008-355-25 (1-10) x AAQ27987 (1-1586)
QY      1 ThrG1G1yASnSerGlySerProValPhe 10
        |||||
DB      1048 ACTGCTGCTACTCAGGTTCCACCGATATT 1077

RESULT 4
AA143635
ID      AA143635 standard; DNA; 2139 BP.
XX
XX      AA143635;
XX
XX      05-SEP-2002 (first entry)
XX
XX      Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.
XX
XX      Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;
XX      DPP-7 inhibitor identification; periodontal disease; gingivitis;
XX      periodontitis.
XX
XX      Porphyromonas gingivalis.
XX
XX
XX      Key      Location/Qualifiers
XX      CDS      1..2139
XX      FT      /*tag= a
XX      FT      /product= "Porphyromonas gingivalis DPP-7"
XX
XX      WO200238742-A2.
XX
XX      16-MAY-2002.
XX
XX      08-NOV-2001; 2001WO-US46782.
XX
XX      08-NOV-2000; 2000US-246827P.
XX
XX      (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX
```

```
XX
XX      Travis J, Potempa JS, Baudula A, Bugno M;
XX
XX      WPI; 2002-490075/52.
XX      P-PSDB; AA015205.
XX
XX      Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
XX      the dipeptidylpeptidase for protecting an animal from periodontal
XX      disease caused by Porphyromonas gingivalis -
XX
XX      Claim 11; Fig 4; 65pp; English.
XX
XX      The invention comprises the amino acid and coding sequence of the
XX      Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
XX      enzyme has amidolytic activity for cleavage of a peptide bond between the
XX      second and third amino acids from the N-terminal end of a target peptide.
XX      The DPP-7 target peptide has an aliphatic or aromatic residue as a
XX      substituent on the alpha-carbon atom of the second amino acid from the
XX      C N-terminal end. The DPP-7 protein and DNA sequences of the invention are
XX      useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
XX      reducing the growth of a bacterium and protecting an animal from a
XX      periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
XX      or periodontitis). The present DNA sequence encodes the Porphyromonas
XX      gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the invention.
XX
SQ      Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other;

Alignment Scores:
Pred. No.:      3.73      Length:      2139
Score:          54.00      Matches:      10
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:            24      Gaps:      0

US-10-008-355-25 (1-10) x AA143635 (1-2139)
QY      1 ThrG1G1yASnSerGlySerProValPhe 10
        |||||
DB      1930 ACGGGCGTACTCCGGTACGCCCGATTTC 1959

RESULT 5
AAH52485
ID      AAH52485 standard; DNA; 849 BP.
XX
XX      AAH52485;
XX
XX      03-SEP-2001 (first entry)
XX
XX      S. epidermidis open reading frame nucleotide sequence SEQ ID NO:363.
XX
XX      Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX      vaccination; endocarditis; ds.
XX
XX      Staphylococcus epidermidis.
XX
XX      WO200134809-A2.
XX
XX      17-MAY-2001.
XX
XX      09-NOV-2000; 2000WO-US30782.
XX
XX      09-NOV-1999; 99US-0164258.
XX
XX      (GLAX ) GLAXO GROUP LTD.
XX
XX      Kimerly WJ;
XX
XX      WPI; 2001-316495/33.
XX      P-PSDB; AAG81635.
XX
XX      Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX      useful for vaccinating against infections, e.g. endocarditis -
XX
XX      PT
```


CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
SQ Sequence 3189 BP; 1081 A; 596 C; 502 G; 1010 T; 0 other;

Alignment Scores:
Pred. No.: 46.9 Length: 3189
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 22 Gaps: 0

US-10-008-355-25 (1-10) x AAH54330 (1-3189)

OY 2 GLYGLYAsnSerGlySerProValphe 10
ID AAL17622 standard; cDNA; 402 BP.
XX
DB 1147 GGTGGAACCTCTGATCTCCAGTATTT 1121

RESULT 8
AAL17622

AC AAL17622;
XX
DT 07-DEC-2001 (first entry)
XX

DE Human breast cancer expressed polynucleotide 10079.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX

OS Homo sapiens.
XX

PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX

PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer
XX
PS Claim 1; Page 1799; 3695pp; English.
XX

CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL16789) and methods of assessing whether a patient is the
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and

CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.
XX

SQ Sequence 402 BP; 101 A; 98 C; 104 G; 99 T; 0 other;

Alignment Scores:
Pred. No.: 38.4 Length: 402
Score: 44.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 81.48% Indels: 0
DB: 22 Gaps: 0

US-10-008-355-25 (1-10) x AAL17622 (1-402)

OY 1 ThrGLYGLYAsnSerGlySerProValphe 10
ID AAL00017 standard; cDNA; 423 BP.
XX
DB 294 ACCGGGGGAGTTGGGAGACCACTTTT 323

RESULT 9
AAL00017

AC AAL00017;
XX
DT 21-NOV-2001 (first entry)
XX

DE Human reproductive system related antigen cDNA SEQ ID NO: 18.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX
KW cancer; gene therapy; ss.
XX

OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX

PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-020515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
XX

PR	22-AUG-2000	2000US-0226568
PR	22-AUG-2000	2000US-0227182
PR	23-AUG-2000	2000US-0227709
PR	30-AGO-2000	2000US-0228924
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229287
PR	01-SEP-2000	2000US-0229343
PR	01-SEP-2000	2000US-0229345
PR	03-SEP-2000	2000US-0229513
PR	05-SEP-2000	2000US-0229513
PR	06-SEP-2000	2000US-0230437
PR	06-SEP-2000	2000US-0230438
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0231414
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-0232080
PR	12-SEP-2000	2000US-0231968
PR	14-SEP-2000	2000US-0232337
PR	14-SEP-2000	2000US-0232338
PR	14-SEP-2000	2000US-0232398
PR	14-SEP-2000	2000US-0232399
PR	14-SEP-2000	2000US-0232400
PR	14-SEP-2000	2000US-0232401
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233064
PR	14-SEP-2000	2000US-0233065
PR	21-SEP-2000	2000US-0234222
PR	21-SEP-2000	2000US-0234274
PR	21-SEP-2000	2000US-0234274
PR	22-SEP-2000	2000US-0234997
PR	22-SEP-2000	2000US-0234998
PR	26-SEP-2000	2000US-0235484
PR	27-SEP-2000	2000US-0235834
PR	27-SEP-2000	2000US-0235836
PR	27-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236367
PR	29-SEP-2000	2000US-0236369
PR	29-SEP-2000	2000US-0236370
PR	02-OCT-2000	2000US-0236802
PR	02-OCT-2000	2000US-0237037
PR	02-OCT-2000	2000US-0237038
PR	02-OCT-2000	2000US-0237039
PR	02-OCT-2000	2000US-0241185
PR	13-OCT-2000	2000US-0239945
PR	13-OCT-2000	2000US-0239937
PR	13-OCT-2000	2000US-0241807
PR	20-OCT-2000	2000US-0241808
PR	20-OCT-2000	2000US-0241809
PR	20-OCT-2000	2000US-0241816
PR	20-OCT-2000	2000US-0241817
PR	01-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246474
PR	08-NOV-2000	2000US-0246475
PR	08-NOV-2000	2000US-0246476
PR	08-NOV-2000	2000US-0246477
PR	08-NOV-2000	2000US-0246478
PR	08-NOV-2000	2000US-0246523
PR	08-NOV-2000	2000US-0246524
PR	08-NOV-2000	2000US-0246525
PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	17-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0246617
PR	17-NOV-2000	2000US-0246907

PR	17-NOV-2000;	2000US-0249208.	
PR	17-NOV-2000;	2000US-0249209.	
PR	17-NOV-2000;	2000US-0249210.	
PR	17-NOV-2000;	2000US-0249211.	
PR	17-NOV-2000;	2000US-0249212.	
PR	17-NOV-2000;	2000US-0249213.	
PR	17-NOV-2000;	2000US-0249214.	
PR	17-NOV-2000;	2000US-0249215.	
PR	17-NOV-2000;	2000US-0249216.	
PR	17-NOV-2000;	2000US-0249217.	
PR	17-NOV-2000;	2000US-0249218.	
PR	17-NOV-2000;	2000US-0249219.	
PR	17-NOV-2000;	2000US-0249220.	
PR	17-NOV-2000;	2000US-0249221.	
PR	17-NOV-2000;	2000US-0249222.	
PR	17-NOV-2000;	2000US-0249223.	
PR	17-NOV-2000;	2000US-0249224.	
PR	17-NOV-2000;	2000US-0249225.	
PR	17-NOV-2000;	2000US-0249226.	
PR	17-NOV-2000;	2000US-0249227.	
PR	17-NOV-2000;	2000US-0249228.	
PR	17-NOV-2000;	2000US-0249229.	
PR	17-NOV-2000;	2000US-0249230.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251088.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251477.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251866.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	08-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX			
PA	(HUMA-)	HUMAN GENOME SCI INC.	
XX			
PI	Rosen CA,	Barash SC,	Ruben SM;
XX			
DR	WPI:	2001-465570/50.	
DR	P-PSDB:	AAM94047.	
XX			
PT	Isolated nucleic acid molecule encoding a reproductive system antigen -		
PT	is used in preventing, treating or ameliorating a medical condition -		
XX			
XX			
XX	Claim 1;	SEQ ID NO 18;	1297pp + Sequence Listing; English.
XX			
XX			
CC	The present invention provides the protein and coding sequences of a		
CC	number of human reproductive system related antigens. These can be used		
CC	in the prevention and treatment of reproductive system disorders,		
CC	including cancer. The present sequence is a coding sequence of the		
CC	invention.		
XX			
SQ	Sequence	423 BP;	93 A; 114 C; 109 G; 104 T; 3 other;
	Alignment Scores:		
	Pred. No.:	40.6	Length: 423
	Score:	44.00	Matches: 7
	Percent Similarity:	90.00%	Conservative: 2
	Best Local Similarity:	70.00%	Mismatches: 1
	Query Match:	81.48%	Indels: 0
	DB:	22	Gaps: 0
US-10-008-355-25 (1-10) x AAL00017 (1-423)			
OY	1	ThrglyGlyAsnserGlySerProvalPhe	10
		::: :::	
Db	300	ACGGCGGAGAAATGCGAGTCCATCTTC	329
RESULT 10			
ID	AAX20213		
XX	AAX20213;		
AC			
XX			
PT	20-Apr-1999 (first entry)		

XX Enterococcus faecalis EF110 gene fragment.
 DE
 XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KW detection; attenuation; antigenic; ss.
 XX
 XX Enterococcus faecalis.
 OS
 PN WO9850554-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08959.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
 XX
 DR WPI: 1999-070095/06.
 DR P-PSDB; AAY0223.
 XX
 PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection
 XX
 PS Claim 1; Page 212-213; 301pp; English.
 XX
 CC The present sequence encodes an antigenic polypeptide fragment
 CC isolated from Enterococcus faecalis. The present invention describes
 CC genes, proteins and antigenic polypeptides isolated from E. faecalis.
 CC The proteins can be used in vaccines for preventing or attenuating an
 CC infection caused by a member of the Enterococcus genus in an animal.
 CC They can also be used for detecting Enterococcus antibodies in a sample.
 CC The nucleotide sequences can be used for detecting Enterococcus nucleic
 CC acids. Products from the present invention can also be used for
 CC screening compounds to identify agonists and antagonists of E. faecalis
 CC protein activity.
 XX
 SQ Sequence 758 BP; 248 A; 143 C; 160 G; 207 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 77.5 Length: 758
 Score: 44.00 Matches: 7
 Percent Similarity: 90.00% Conservative: 2
 Best Local Similarity: 70.00% Mismatches: 1
 Query Match: 81.48% Indels: 0
 DB: 20 Gaps: 0
 US-10-008-355-25 (1-10) x AAX20213 (1-758)
 OY 1 ThrGlyGlyAsnSerGlySerProValphe 10
 ||||||| |||||||
 DB 575 ACCGGCGTCATCTGTTCCACCATCTAT 604
 RESULT 11
 ABN98198
 ID ABN98198 standard; DNA; 758 BP.
 XX
 AC ABN98198;
 XX
 DT 05-AUG-2002 (first entry)
 XX
 DE E faecalis EF110 gene fragment.
 DE
 KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
 KW gene; ds.
 XX
 PA Enterococcus faecalis.
 OS
 PI

PN US2002045737-A1.
 XX
 PD 18-APR-2002.
 XX
 PF 04-MAY-1998; 98US-0071035.
 XX
 PR 04-MAY-1998; 98US-0071035.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;
 XX
 DR WPI: 2002-425450/45.
 DR P-PSDB; ABP43442.
 XX
 PT New genes and polypeptides from Enterococcus faecalis, useful as
 PT vaccines for preventing, treating or attenuating an infection caused by
 PT a member of the Enterococcus genus in an animal, particularly E.
 PT faecalis -
 XX
 PS Claim 1; Page 197-198; 255pp; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of polypeptides from Enterococcus faecalis. The proteins can be
 CC used as vaccines for preventing or attenuating an infection caused by a
 CC member of the Enterococcus genus in an animal, particularly E. faecalis.
 CC The polynucleotide is also useful for preventing or treating E. faecalis
 CC infection. The present sequence is a coding sequence of the invention.
 XX
 SQ Sequence 758 BP; 248 A; 143 C; 160 G; 207 T; 0 other;
 XX
 Alignment Scores:
 Pred. No.: 77.5 Length: 758
 Score: 44.00 Matches: 7
 Percent Similarity: 90.00% Conservative: 2
 Best Local Similarity: 70.00% Mismatches: 1
 Query Match: 81.48% Indels: 0
 DB: 24 Gaps: 0
 US-10-008-355-25 (1-10) x ABN98198 (1-758)
 OY 1 ThrGlyGlyAsnSerGlySerProValphe 10
 ||||||| |||||||
 DB 575 ACCGGCGTCATCTGTTCCACCATCTAT 604
 RESULT 12
 AAX20212
 ID AAX20212 standard; DNA; 888 BP.
 XX
 AC AAX20212;
 XX
 DT 20-APR-1999 (first entry)
 XX
 DE Enterococcus faecalis gene EF110.
 DE
 KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KW detection; attenuation; antigenic; ss.
 XX
 OS Enterococcus faecalis.
 OS
 PN WO9850554-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08959.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;

XX WPI: 1999-070095/06.
DR P-PSDB: AAY00222.
XX
PT New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection
XX
PS Claim 1: Page 212; 301pp; English.
XX
CC The present sequence represents a gene isolated from
CC Enterococcus faecalis. The present invention describes genes, proteins
CC and antigenic polypeptides isolated from E. faecalis. The proteins can
CC be used in vaccines for preventing or attenuating an infection caused
CC by a member of the Enterococcus genus in an animal. They can also be
CC used for detecting Enterococcus antibodies in a sample. The nucleotide
CC sequences can be used for detecting Enterococcus nucleic acids.
CC Products from the present invention can also be used for screening
CC compounds to identify agonists and antagonists of E. faecalis protein
CC activity.
XX
SQ Sequence 888 BP; 290 A; 159 C; 186 G; 253 T; 0 other;
XX
Alignment Scores:
Pred. No.: 92.3 Length: 888
Score: 44.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 81.48% Indels: 0
DB: 20 Gaps: 0
US-10-008-355-25 (1-10) x AAY20212 (1-888)
QY 1 ThrGlyGlyAsnSerGlySerProValphe 10
Db 703 ACCGGCGTCATCTGTTCCACCAATCTAT 732
RESULT 13
ABN98197
ID ABN98197 standard: DNA; 888 BP.
XX
AC ABN98197;
DT 05-AUG-2002 (first entry)
XX
DE E faecalis EF110 gene.
XX
KM Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
KM gene; ds.
XX
OS Enterococcus faecalis.
XX
PN US2002045737-A1.
PD 18-APR-2002.
XX
PF 04-MAY-1998; 98US-0071035.
XX
PR 04-MAY-1998; 98US-0071035.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PI Chol GH, Bailey C, Hromockyj A, Kunsch CA;
XX
DR WPI: 2002-425450/45.
DR P-PSDB: ABP43441.
XX
PT New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly E.
PT faecalis -
PS Claim 1: Page 197; 255pp; English.

XX
CC The present invention provides the protein and coding sequences of a
CC number of polypeptides from Enterococcus faecalis. The proteins can be
CC used as vaccines for preventing or attenuating an infection caused by a
CC member of the Enterococcus genus in an animal, particularly E. faecalis.
CC The polynucleotide is also useful for preventing or treating E. faecalis
CC infection. The present sequence is a coding sequence of the invention.
XX
SQ Sequence 888 BP; 290 A; 159 C; 186 G; 253 T; 0 other;
XX
Alignment Scores:
Pred. No.: 92.3 Length: 888
Score: 44.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 81.48% Indels: 0
DB: 24 Gaps: 0
US-10-008-355-25 (1-10) x ABN98197 (1-888)
QY 1 ThrGlyGlyAsnSerGlySerProValphe 10
Db 703 ACCGGCGTCATCTGTTCCACCAATCTAT 732
RESULT 14
AAV80633
ID AAV80633 standard: CDNA; 2421 BP.
XX
AC AAV80633;
DT 01-MAR-1999 (first entry)
XX
DE Kidney injury associated molecule HW096 CDNA clone.
XX
KM Kidney injury associated molecule; kidney injury related molecule;
KM KIM; tissue growth promotion; regeneration; renal condition;
KM acute renal failure; acute nephritis; tumour; ds.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT CDS 295..1305
FT /*tag=a
FT /product="kidney injury associated molecule"
XX
PN W09853071-A1.
XX
PD 26-NOV-1998.
XX
PF 22-MAY-1998; 98WO-US10547.
XX
PR 23-MAY-1997; 97US-0047491.
XX
PR 23-MAY-1997; 97US-0047490.
XX
PA (BIOJ) BIOGEN INC.
PI Cate RL, Hession CA, Sanicola-Nedel M, Wei H;
XX
DR WPI: 1999-045312/04.
DR P-PSDB: AAW86338.
XX
PT Kidney injury-associated molecule, KIM, polypeptides - upregulated
PT in injured or regenerating tissues, useful to promote tissue growth
PT and regeneration, especially to treat renal conditions
XX
PS Claim 9: Page 186-188; 213pp; English.
XX
CC The present sequence represents a kidney injury associated molecule
CC (KIM) CDNA clone. KIM proteins can be administered therapeutically
CC by expressing KIM encoding polynucleotides, to promote growth and/or
CC survival of damaged tissue (e.g. renal tissue), since the KIM proteins
CC are upregulated in injured or regenerating (especially renal) tissues.
CC KIM fusion proteins, conjugates, antibodies and vectors can also be used

CC therapeutically, e.g. these or the KIM proteins may be included with an
CC acceptable carrier in pharmaceutical compositions, useful for therapy/
CC prophylaxis of conditions associated with dysfunction/dysregulation of
CC KIM genes or proteins, especially renal diseases or impairments of renal
CC function in humans (e.g. acute renal failure, acute nephritis). The
CC polynucleotides can be used to produce antisense sequences which, when
CC internalised into cells, can disrupt expression of a cellular KIM gene,
CC also useful in therapy (e.g. to block the growth of tumours dependent on
CC KIM for growth) or compositions. The proteins and polynucleotides are
CC useful diagnostically e.g. to detect and quantify renal injury/disease
CC (indicative of increased risk, or presence of, renal injury or impaired
CC function), or abnormal responses to tissue injury (indicative of
CC increased risk, or presence of, an autoimmune response or abnormal
CC tissue growth arising from/affecting renal tissue). The proteins can
CC also be used to locate KIM-producing cells (especially specific loci,
CC e.g. tissue masses abnormally producing/expressing KIM such as tumours
CC arising from/affecting renal tissue), by contacting cells with an
CC imageable KIM-binding reagent and imaging reagent accumulation.

XX Sequence 2421 BP; 785 A; 417 C; 533 G; 686 T; 0 other;

Alignment Scores:

Pred. No.:	280	Length:	2421
Score:	44.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	88.89%	Mismatches:	0
Query Match:	81.48%	Indels:	0
DB:	20	Gaps:	0

US-10-008-355-25 (1-10) x AAV80633 (1-2421)

QY 2 GLYGLYAsnSerGlySerProValPhe 10

Db 1087 GGTGGGCTCTCTGATGCCACGATATT 1113

RESULT 15

AAAX13198/C
ID AAX13198 standard; DNA: 4951 BP.

XX AAX13198;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:261.

XX Enterococcus faecalis; contig; detection; Enterococcal infection;

KM vaccine; attenuation; computer readable medium; ds.

XX Enterococcus faecalis.

OS Enterococcus faecalis.

XX Enterococcus faecalis.

PN MO9850555-A2.

XX MO9850555-A2.

PD 12-NOV-1998.

XX 12-NOV-1998.

PF 04-MAY-1998; 98MO-US08985.

XX 04-MAY-1998; 98MO-US08985.

PR 14-NOV-1997; 97US-0066009.

XX 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

XX 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX 16-MAY-1997; 97US-0046655.

CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.

XX Sequence 4951 BP; 1563 A; 981 C; 735 G; 1665 T; 7 other;

Alignment Scores:

Pred. No.:	618	Length:	4951
Score:	44.00	Matches:	7
Percent Similarity:	90.00%	Conservative:	2
Best Local Similarity:	70.00%	Mismatches:	1
Query Match:	81.48%	Indels:	0
DB:	20	Gaps:	0

US-10-008-355-25 (1-10) x AAX13198 (1-4951)

QY 1 ThrGLYGLYAsnSerGlySerProValPhe 10

Db 695 ACCGCGCTCAATCTGCTTCACCAATCTAT 666

Search completed: December 20, 2002, 13:22:50
Job time : 197.526 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:11:39 ; Search time 1514.74 Seconds

(Without alignments)
192.131 Million cell updates/sec

Title: US-10-008-355-25
Perfect score: 54
Sequence: 1 TCGNSGSPVF 10

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n model -DEV=vlh
-O/cg22_1/USPTO.spool/US10008355/runat_17122002_112337_14593/app_query.fasta.1.398
-DB=GenBml -OFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USRR=US10008355 @CCN_1.1.1616 @runat_17122002_112337_14593 -NCRU=6 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenBml:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	1289	1 STAGASP	D00730 S. aureus g
2	54	100.0	1558	6 E03836	E03836 DNA encodin
3	54	100.0	1586	6 E03835	E03835 DNA encodin
4	54	100.0	1634	1 SASP	V00356 Staphylococ
5	54	100.0	3240	1 AF309515	AF309515 Staphyloc
6	54	100.0	5207	1 SWA29385	SWA29385 Staphyloc
7	54	100.0	290150	1 AP004825	AP004825 Staphyloc
8	54	100.0	298050	1 AP003132	AP003132 Staphyloc
9	54	100.0	347235	1 AP003361	AP003361 Staphyloc
10	49	90.7	657	1 SEP305145	AX305145 Staphyloc
11	49	90.7	849	6 AX141641	AX141641 Sequence
12	49	90.7	3189	1 AF269652	AF269652 Staphyloc
13	49	90.7	3189	6 AX144972	AX144972 Sequence
14	48	88.9	10689	1 AE004008	AE004008 Xylella f
15	45	83.3	66762	2 AC121125	AC121125 Mus muscu
16	45	83.3	120355	9 AC011338	AC011338 Homo sapi
17	45	83.3	175191	9 AC010251	AC010251 Homo sapi
18	45	83.3	194355	2 AC116511	AC116511 Mus muscu
19	45	83.3	195335	2 AC068065	AC068065 Mus muscu
20	45	83.3	260050	1 AL596166	AL596166 Listeria
21	45	83.3	349980	6 AX417039	AX417039 Sequence
22	45	83.3	349980	6 AX417042	AX417042 Sequence
23	44	81.5	1085	1 EFSPREG	Z12286 E. faecalis
24	44	81.5	1912	9 AF008576	AF008576 Homo sapi
25	44	81.5	2385	10 BC019638	BC019638 Mus muscu
26	44	81.5	36547	2 AC017307	AC017307 Drosophila
27	44	81.5	40221	1 MSGY154	AB000002 Mycobacte
28	44	81.5	58722	2 AC100192	AC100192 Mus muscu
29	44	81.5	110000	2 AC119565-2	Continuation (3 of
30	44	81.5	162609	3 AC108022	AC108022 Homo sapi
31	44	81.5	166863	3 AC009741	AC009741 Drosophila
32	44	81.5	166978	2 AC108085	AC108085 Homo sapi
33	44	81.5	171279	2 AC036185	AC036185 Homo sapi
34	44	81.5	174253	2 AC068590	AC068590 Homo sapi
35	44	81.5	178199	3 AC007724	AC007724 Homo sapi
36	44	81.5	182036	9 AC073533	AC073533 Homo sapi
37	44	81.5	187036	2 AC124514	AC124514 Homo sapi
38	44	81.5	189014	2 AC051649	AC051649 Homo sapi
39	44	81.5	193946	2 AC121339	AC121339 Homo sapi
40	44	81.5	206924	2 AL845498	AL845498 Mus muscu
41	44	81.5	222941	2 AC102236	AC102236 Mus muscu
42	44	81.5	225635	3 AE003695	AE003695 Drosophila
43	44	81.5	234431	9 AF311103	AF311103 Homo sapi
44	44	81.5	279166	2 AC126675	AC126675 Mus muscu
45	44	81.5	346897	1 AP002995	AP002995 Mesorhizo

RESULT 1

ALIGNMENTS

[illegible]

```

REFERENCE          Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS            I (bases 1 to 1558)
                   Nakamura,E., Yamaki,M., Teraoka,H., Matsumoto,K., Shin,M.,
                   Fujiwara,K., Tsuzuki,H., Yoshida,N. and Kakudou,S.
TITLE              NEW PROTEASE
JOURNAL            Patent: JP 1992211370-A 2 03-AUG-1992;
                   SHIONOGI & CO LTD
COMMENT            OS   Staphylococcus aureus
                   PN   JP 1992211370-A/2
                   PD   03-AUG-1992
                   PF   19-FEB-1991 JP 1991024633
                   PR   20-FEB-1990 JP 90P 40398
                   PI   NAKAMURA ETSUO, YAMAKI MIKIO, TERAOKA HIROSHI, PI MATSUMOTO
                   KOCHI,
                   PI   SHIN MASARU, FUJIWARA KOJI, TSUZUKI HIROSHIGE, YOSHIDA NOBUO,
                   PI KAKUDOU SHINJI
                   PC   C12N9/53,C12N15/57,(C12N15/57,C12R1:445);
                   CC   strandedness: Double;
                   CC   topology: Linear;
                   CC   hypothetical: No;
                   CC   anti-sense: No;
                   CC   *source: strain=V8;
                   CC   *source: clone=PAM82S1GVB;
                   CC   Feature is identified by similarity;
                   FH   key Location/Qualifiers
                   FT   CDS             354..1364
                   FT     /product='precursor of V8 protease' FT
                   FT   sig_peptide    354..557
                   FT     mat_peptide   558..1361
                   FT     /product='V8 protease'
FEATURES           source
                   Location/Qualifiers
                   1..1558
                   /organism="Staphylococcus aureus"
                   /db_xref="taxon:1280"
BASE COUNT         579 a      282 c      221 g      476 t
ORIGIN
Alignment Scores:
Pred. No.:        18.7          Length:       1558
Score:            54.00         Matches:      10
Percent Similarity: 100.00%     Conservative: 0
Best Local Similarity: 100.00% Mismatches:    0
Query Match:      100.00%       Indels:       0
DB:               Gaps:         0
US-10-008-355-25 (1-10) x E03836 (1-1558)
QY                1 ThrglgyIyaSnserglySerProValphe 10
                  |||||||.....
Db 1050 ACTGGTGTGAATTCAGTTCACTCATATTATT 1079
RESULT 3
LOCUS              E03835                      1586 bp      DNA      linear      PAT 29-SEP-1997
DEFINITION         DNA encoding V8-like protease.
ACCESSION          E03835
VERSION            E03835.1 GI:2172049
KEYWORDS           JP 1992211370-A/1.
SOURCE             Staphylococcus aureus.
ORGANISM           Staphylococcus aureus
REFERENCE           Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS            1 (bases 1 to 1586)
                   Nakamura,E., Yamaki,M., Teraoka,H., Matsumoto,K., Shin,M.,
                   Fujiwara,K., Tsuzuki,H., Yoshida,N. and Kakudou,S.
TITLE              NEW PROTEASE
JOURNAL            Patent: JP 1992211370-A 1 03-AUG-1992;
                   SHIONOGI & CO LTD
COMMENT            OS   Staphylococcus aureus
                   PN   JP 1992211370-A/1
                   PD   03-AUG-1992
                   PF   19-FEB-1991 JP 1991024633
                   PR   20-FEB-1990 JP 90P 40398

```



```
/translation="MKGRFLKYSLFVATLTATLVSSPANALSSKAMNDHPQOTS
SKQOTPKLOKGNLKEPDRERHANYILPNNDHQTIDTNGHYATVITYOVARPTGE
IASGVVGVGKDLTLTKMKHYVDAGDPDLAKAPPSA1INDONNGEFTLEOLTKYSGEG
DLAIYKFSRPNENKHIIGEVKRAIAMSNAETOVNONITYTGTGPKRPAITMESGKI
TYLKGEAAQIDLTSTGGNSGSEVENKEKEVIGIHGSGVPENEGAVFIENEENRFLKO
NIEDIHFNADDPNNPNDPNPNNDPNNDPNNDPNNDPNNDPNNDPNNDPNNDPNND
AA"
1446..2627
/gene="sspb"
1446..2627
/gene="sspb"
/feature="sspb"
/feature="sspb: preproprotein"
/codon_start=1
/product="cysteine protease precursor"
/protein_id="AAG45844.1"
/db_xref="GI:12025239"
/translation="MNSCKSRFVNIISIINVSMLISLGAFAANNKAKADSHSKOLE
INVKSDKVPQKYLDAQOOFAGYAKALDKQSAKGEKELGEAFKIFKENGEDNSY
YVIVKDKIVYTLTSPKNDLNSKEDMNSYKISNFIADLDQIKDKNSNITVL
DEKGFYFEDDKVRLVKATPLPGNVKESAKTSAKIKOELKNTVTPTKYVENALIO
EDQVYENTLAKFKIREQGFDSNWCAGFSMALMLATKNTDTTNAHDIRITIPRESE
ODLPNCATFPMQMIETGKSQGRDHIYQGVPSYEQVDLTQNVGIMTLAOSVQNP
DPHLGHALAVGNAKINDQEKLIYMPDTELSIODADSLHLHSFNRDYMWYGSIG
Y"
1446..1553
/gene="sspb"
2103..2624
/gene="sspb"
/product="cysteine protease"
1554..2624
/gene="sspb"
/feature="sspb"
/feature="sspb: encodes the cysteine protease proprotein"
2665..2994
/gene="sspc"
2665..2994
/gene="sspc"
/feature="sspc"
/feature="sspc: unknown: possesses no signal sequence: predicted
cytoplasmic location"
/codon_start=1
/translation="1"
/product="1"
/protein_id="AAG45845.1"
/db_xref="GI:12025240"
/translation="MYOLOFINLVYDTKTLHLEQTNINIFGNNSNHOLOKSTCIRH
GDDPSHNOYHILFIDTAHORIKFSSIDDEIITIIIDYDTQHIIMQTSKOSTISRP
IYERLV"
BASE COUNT 1254 a 547 c 461 g 978 t
ORIGIN
Alignment Scores:
Pred. No.: 32 Length: 3240
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-10-008-355-25 (1-10) x AF309515 (1-3240)
QY 1 ThrGlyGlyAsnserGlyserProvalphe 10
|||||
Db 1050 ACTGGTGTAATTCAAGGTCACCTGATTTT 1079
RESULT 6
SMA293885 5207 bp DNA linear BCT 12-AUG-2002
LOCUS Staphylococcus warneri prom gene, proc gene, prod gene, spw1 gene
DEFINITION
ACCESSION AJ293885
VERSION AJ293885.2 GI:22218025
KEYWORDS Cysteine proteinase; glutamyl endopeptidase; msw1 gene; proc gene;
prod gene; prom gene; spw1 gene.
```

```
SOURCE
ORGANISM Staphylococcus warneri.
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1 Yokoi,K., Kakikawa,M., Kimoto,H., Watanabe,K., Yasukawa,H.,
Yamakawa,A., Taketo,A. and Kodaira,K.I.
Genetic and biochemical characterization of glutamyl endopeptidase
of Staphylococcus warneri M
Gene 281 (1-2), 115-122 (2001)
2 21623048
JOURNAL
MEDLINE
REFERENCE
Kakikawa,M.
Direct Submission
Submitted (30-AUG-2000) Kakikawa,M., Molecular Biology Group,
Toyama University, 3190 Gofuku, Toyama, 930-8555, JAPAN
REMARK
COMMENT
On Aug 13, 2002 this sequence version replaced gi:9968802.
FEATURES
source
1..5207
/organism="Staphylococcus warneri"
/db_xref="taxon:1292"
complement(232..708)
/gene="msw1"
complement(232..708)
/gene="msw1"
/codon_start=1
/translation="1"
/protein_id="CAD43735.1"
/db_xref="GI:22218026"
/translation="MEIYVAGGCLMGVQAEFFKIPGIIVTEAGRANGITSNLGEYD
GVAECVKLTDPDLSLTIRNMSYLFEPVSDPILNOGODDICEKRTGTGYSNTKHLNE
AOTFEREDRAHIVEVLPISLYKSPBHQDRLENPEBHCHLITMHLNKKY"
991..2148
/gene="spw1"
991..2148
/gene="spw1"
/codon_start=1
/translation="1"
/protein_id="CAD43736.1"
/db_xref="GI:22218027"
/translation="MTHSIKYLLGATITTSALFLTSHHEASATIENTNOVDGFRNQ
SNGNALITQANOTDSTYNNHDENGNPJAVNNIINDNFNHNGDAYSDDNVATFNHDE
AGNPLMTANSVDSDFNHDONGNPJAEKNIDSDFNHDONGNPJAVNNIINDNFNHNG
GDAIYSDNNVDSAEFNHDEGPNPLMTANSVDSDFNHDONGNPJASNNIINDNFNHNG
DAISDDTDAFNFHDEGNPLMTANSVDSDFNHDONGNPJASNNIINDNFNHNGDGN
FTASDNNIINDNFNHNGEATVANKNAVDGPNHDAVGPJIANNOAMITSRMOOSHDT
TTPOSATITTHOQNSLPATGODRRAPYIGVITLVLAFAFLLRKSSAK"
2432..3382
/gene="prom"
2432..3382
/gene="prom"
/feature="prom"
/codon_start=1
/translation="1"
/protein_id="glutamyl endopeptidase"
/db_xref="GI:9968803"
/db_xref="SPTREMBL:Q9PBG1"
/translation="MKVKEFTASSLLILTLTSAIILNPAHETTSIDNHQOTTQSQ
QKTPKIDKGNVNVKPEKKERANVILPNNDRIQINDTTLGHAAPVTFVOVQNEGFIA
SGVVVGKDKLTLTKMKHYVDATGPNPALAPPSAVQNNY PNGGFTGEQITKYPGNAKL
AIVKFSRPNONONIGEVTPATLSNADTSQNPITVYGYGDKPLATMWSRSKTIIO
IOGEDMHYDLSSTGGNSGSPVENSRENEVIGIHGGAANSYGAVFINNVGNFLKONI
EDIHFSNSDNDNDNDNNCTDDNNNNNDNDNDNDNDNDNDNDNDNDNDNDNDNDND
2620..3378
/gene="prom"
/feature="prom"
/product="glutamyl endopeptidase"
/evidence="experimental"
3471..4661
/gene="proc"
3471..4661
/gene="proc"
/feature="proc"
/codon_start=1
/translation="1"
mat_peptide
gene
CDS
```

/product="cysteine proteinase"
 /protein_id="CAD43737.1"
 /db_xref="GI:22218028"
 /translation="MHASYKHISFKLISLILFAILLSTIAFDLIDKNSQNSOLE
 INKNDVPORVONLAQQOYKGYAQAALDKRNSQTCGYLGAEKIRKFGNEENSYY
 YPKADKRLIYTLTISPRNEADLKOSKENNYSKISNPLAKDLDEIKDHTNITLT
 DEYGLFEENGQOYRLVAKAPLPTNKKKKNSSOHSSHLKQOLAKATSEPTINENQAE
 ANDNOQYENLTLNKFKIRRCQCPDNSCAFSMAALLNATKNTDKYNAHDIRMLTYR
 VSEELPQESTNPHQMTKYNSSQGRDLQHSQGMTYQTDQLTKDNGAMILIASVSN
 NPNDPLGHAMAVVGNKINNEDKLIFNPWDITLSTQDADSNLLHLSFNDYTWDDT
 MGY"
 4698 .5024
 /gene="5024"
 /gene="prod"
 4698 .5024
 /gene="prod"
 /codon_start=1
 /transl_table=1
 /protein_id="CAD43738.1"
 /db_xref="GI:22218029"
 /translation="MYOLFRIHINDALTLTKSKODTIHLFIGWNIWNSAKSISIRP
 GYDHNHQYOIQLDITDHQRIKLISEEDPOLMTILDYEDINHIIFQTSYKNSYGTSSRP
 IRIEKF"
 1937 a 988 c 751 g 1531 t

gene 285..1493
/gene="MW0823"
CDS 285..1493
/gene="MW0823"
/note="ORFID:MW0823
hypotheetical protein, similar to NADH dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB94688.1"
/db_xref="GI:21203990"
/translation="MAQDRKRVLVUGAGYAGLTQVTQLQKAISTEEAEITLTKNKEYH
YEATLVHASAGTLNVEDLVYVESVYLAKDKRVNFAQEVKRTIDDAKRVFTNGITDPL
DILVVALFVSEETGIEGMRKHAFQIENVITARELSRIIEDEKFNPAASKEKDNIDSL
ILVAGAGTGEVFLGELTDRIPELCSKYGVONKVKITCYEAAKMLPMSEELVYNIA
VSYLDRQVREIAPPIVACNKGKGVNDEKQOLNAGTSMVAAVGKSGKLMSEEP
GKKRRIITKDDITNGINDYFNVTGDCSAFLPAEEERDLPPTTAQIMADGGSVAKNK
RLNKGSEPEEPYVDRCIVCSLGHSDGVGMFDPRIDAGKRAAFKKYIDIRAVKICG
IGLAKKKGF"
gene 1624..3099
/gene="ampl"
CDS 1624..3099

Assignment Scores:	
Pred. No.:	45.3
Score:	54.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	1
Length:	5207
Matches:	10
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-008-355-25 (1-10) x SWA293885 (1-5207)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10
| | | | |
DB 3122 ACTGGTGTAAATTCAGGTTCCACCTGTATTTC 3151

RESULT 7
AP004825/C

LOCUS	290150 bp	DNA	linear	BCT 02-JUL-2002
DEFINITION	Staphylococcus aureus subsp. aureus MW2 DNA, complete genome, strain:MW2, section 4/10.			
ACCESSION	Ap004825	BA000033		
VERSION	Ap004825.1	GI:21203989		

SOURCE Staphylococcus aureus subsp. aureus MW2 (strain:MW2) DNA

Bacteria; Firmicutes; Bacillales; Staphylococcus.

REFERENCE
1
AUTHORS
Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A., Nagai, Y., Iwama, N., Asano, K., Naimi, T., Kuroda, H., Cui, L., Yamamoto, K. and Hiramatsu, K.

TITLE	Genome and virulence determinants of high virulence community-acquired MRSA
JOURNAL	Lancet 359 (9320), 1819-1827 (2002)

MEDLINE	22040717
PUBMED	12044378
REFERENCE	2 (bases 1 to 290150)

TITLE Hiroda, M., Hiramatsu, K. and Kituchi, H.
JOURNAL Direct Submission
Submitted (06-MAR-2002) Aiko Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 49-10
Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan
(E-mail: oguchi@nite.go.jp, <http://www.bio.nite.go.jp/>,
Tel: 81-3-3481-8423, Fax: 81-3-3481-8424)
FEATURES Location/Qualifiers

```
source
1. 290150
   /organism="Staphylococcus aureus subsp. aureus MW2"
   /strain="MW2"
   /db_xref="taxon:196620"
```

gene 285. .1493
/gene="MM0823"
285. .1493
/gene="MM0823"
/note="ORFID:MM0823
hypothetical protein, similar to NADH dehydrogenase"
/codon_start=1
/transl_table=1
/protein_id="BAB94688.1"
/db_xref="gi:21203990"
/translation="MAODRKLVLVGAGAGLQIVTKLQKASITREAEITLNNNEH
YEATVLAHASAGTLNIEDVLYPVESYLAKDKVNVQAEVTKTIDDAKAVETQNTDIF
DILVALALFVSETPGIEGMDKHAFOIENVTVARELSRIIEDEKFAVNAASEKQNDIIS
ILVGGATGCGVEFGLFDRIPELCSKGYDQNNKVTTCVEAPKMLPMSEELVNHNA
VSYLSDRVEERKIAPIVACNKGKGVVEDEKQOLNAGTSVMAAGVSGSLIMESEP
GKRRIRITKQDITNGVDNI FVIGDCSAFLPAEEBRLPTTAQIMADGESAANKIK
RLINAESEPEEYDNRGTVCSLGSHDVGWFDKPIAGKKAAPKAYIDTRAAYK IGG
IGLAKRKKPF"
1624. .3099
/gene="ampa"
1624. .3099
/gene="ampa"
/note="ORFID:MM0824"
/codon_start=1
/transl_table=1
/product="probable cytosol aminopeptidase"
/protein_id="BAB94688.1"
/db_xref="gi:21203991"
/translation="MNFRLNNTLSNEINTLLIGIPEHLNQLERISFNHDIITDESLERL
KHQHIGSKVGKGIYTTAFDVOQDQTYRLITTVGLNMLKTRSYQDMKLINGHILFOYLKSEH
IEDPTLDSFKSDIQLDQSLVLMACGIIQSEBATEYDFHFKSKKAPKNTLNISESL
IELDIITHGISIGISINLARDPESNMPAVULLPQTFARDIIVNHFKNPKYKVDKQYDILL
VSEGGLLQANGKSGKHKRPLRYITTYNKKDKDEAPIALVGGGITYDSGGISIKRKNM
ATKMDKCGGAANVVGIIEAASRLQLPNIVGLVLAQAEMLINEASMKPDVDFVALISGTE
VEVMTDGEKAVNLADAVFANVQOVSIVMDFAITLGAATVALGDCKALFAAFENSKSVIT
LNDIIQISEVDEWFEELPIATERASISIKHSDIADLVNHTNGQCKALFAAFSVYHFGSG
QTPRHPIAGPATNKAASYNGPKPTGFMTPIYQWLKQO"
3510. .4826
/gene="MM0825"
3510. .4826

gene
/gene="MM0825"
/note="ORFID:MM0825"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94690.1"
/db_xref="GI:21203992"
/translation="MINAVTAVIIMVLICRLNVTLSPLFALVGLSGMSIEKV
INVEKNTVDGEVALSVALLGGFALISYSGIDYLVGKIINIHENSGMSVKKK
VTIITALLAMSTMSQNLIPVHIATPIYIPILSLSPNDKIDRLIGLITSGICPPY
VLIPGEQFOQLIQSGFAPKAMPITERNIMTKMLIPSMQIVGLLGLVFKRPE
YETRRISQSDVNETLRPYLLVTIVAILATFVLEFDSDMEGLAGVLAFFISRAVN
WELEDAKPEVEGKIMAYIGVILLFANGFAGVMAATGDIIDVKTLSITGQNKLFSLI
MMYVILVTVLGLIGSSPATIPIIASLIFPGFASIGLDTMALIALIGTASALGDGSPA
SOSTGTPAGLVGDGQHDHIRTQCVNPLFYNIPLMIGTAAVVL"
4845 . 5219
/gene="MM0826"
4845 . 5219
/gene="MM0826"
/note="ORFID:MM0826"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94691.1"
/db_xref="GI:21203993"
/translation="MTHLLEFEMSIDHQEDGLVVISNPVTDKVKQFGYLHGASIAI
LGEYFASGSANLIDYTKFPIGLEEMNANHISAKDGHVTATAIHQGKSTHYMDIKI
IKNDEQILTYVGRGTVAIKPLK"
complement(5275 . 6429)
/gene="MM0827"
complement(5275 . 6429)
/gene="MM0827"
/note="ORFID:MM0827"

```
/codon_start=1
/translation="hypothetical protein"
/protein_id="BAB94692.1"
/db_xref="GI:21203994"
/translation="MHWTIGGIQGTIAIAOKLSSGLTDRLLTIIDHEFECCFENS
FENREMPYLSPIVHNVHQPFLKOPAKOHQYNAFYGPORBELTFMFDHIAHAS
KOYLEDCLVOGLVOTLDKQEDKWHIKLEDCOITITDCVIAIGSTINFPMDILDK
ONVNHIEPEKHDOVYDKTDHIVGSGITAAHLAKLHNDMDKTHLMDNDIEHDF
DADPGMLPKMSTFSLSTKSMPERNAIYORERHSGMSHELYLRKKIKNGRIIVHK
TPITDISDGLINTENNSVPTQOINVAITFEQDFMSQPLIKOLIQNDAPINECNFVI
SEKLEWIPNLEVAGCFADLEGPGRNMGKRAERLEQAFLEKLDQYSA"
complement(6678..7034)
/gene="mhg"
/complement(6678..7034)
/gene="mhg"
/note="ORFID:MW0828"
/codon_start=1
/translation="Na+/H+ antiporter subunit"
/protein_id="BAB94693.1"
/db_xref="GI:21203995"
/translation="MKIILISLALFIYICGLISALAIAGLRLIEDVYSRAHAKKA
STLGAMSLLEGFLEFYATQGFVNMQLIAIIFVLITGPLSSHMIMKAAYNIKTPYTK
KTKVDEISEDLDKTKL"
complement(7012..7305)
/gene="mhg"
/complement(7012..7305)
/gene="mhg"
/note="ORFID:MW0829"
/codon_start=1
/translation="Na+/H+ antiporter subunit"
/protein_id="BAB94694.1"
/db_xref="GI:21203996"
/translation="MHNHVIIVIALIIVISMLANLINVVLGSPSLADRVALLDAIGLO
LMAVIALFSLINKYIMIVIMIGIILGTAFAVSKPMDKGVLEHNDHND"
complement(7305..7784)
/gene="mhg"
/complement(7305..7784)
/gene="mhg"
/note="ORFID:MW0830"
/codon_start=1
/translation="Na+/H+ antiporter subunit"
/protein_id="BAB94695.1"
/db_xref="GI:21203997"
/translation="MAVDLVNFIIVAVWLFVNTSYTNNFVLCFIEGLVLYLHRY
LPGRYVITLRIKLVILFLEILKANFDVAKIIKPSIKNEPGEFFVYHDLKDWQ
IVLSNLITTPGIVLVGSDDRKRIYIHAIDFSKEQEVESIKTSLEKIVREGEI"
complement(7786..9282)
/gene="mhg"
/complement(7786..9282)
/gene="mhg"
/note="ORFID:MW0831"
/codon_start=1
/translation="Na+/H+ antiporter subunit"
/protein_id="BAB94696.1"
/db_xref="GI:21203998"
/translation="MIESNMLVLTLPVITAILVFTGKRPIIKRYVALGCTLTTLV
AAIINLVNVHGPRIYELGSMKAPYSIVFADIFSAIITTSIITFAIVLYSQTI
GIERERYYSFVLEMLIGIAGFTTGIDINLNFVEFEELSSYELVIGSTKIOLOE
TIKRYLVNVSSSEFVGVATILSVYGLNLSNKANLSADHSGIVNFTLFIPI
VFATKAGVFPMVLPKSYVAPPIPIITAFGALLTKGVAVIATRLTSFFSDNVSFSH
VYLFALLTITFGCVAAVAVANIKIILVNVMLAVGLVAGVAMTESGMIAGIYYT
LHDMVLKALFLIGIMIKITGADLRQFGLIKRYPLVGSFPAALSGLGIPPLSG
FYGKFFIVOSTFERGFYLSGVILSLSLVLYSVRILOGFEPQPKKYDNNKVDYK
YLTITAIIVAVITVYVGLSADYLVPMKAGAFETFPNPTSYKAVLGSK"
complement(9275..9616)
/gene="mhg"
/complement(9275..9616)
/gene="mhg"
/translation="Na+/H+ antiporter subunit"
/protein_id="BAB94697.1"
/db_xref="GI:21203999"
/translation="MEIIMIFVSNILFAISVLYLSLRIRWGTLLTTHAANFLI
TMGGIKATVPIYEANWKSVDYPIQALILFAIYIAFATTAFLVLAFFRYKELGTDN
VESKGVPEED"
complement(9616..10044)
/gene="mhg"
/complement(9616..10044)
/gene="mhg"
/note="ORFID:MW0833"
/codon_start=1
/translation="Na+/H+ antiporter subunit"
/protein_id="BAB94698.1"
/db_xref="GI:21204000"
/translation="MNRQNDLILQFAVILIFFMVWVGSLPLAGHYTPGCGFVGL
LFASSIVITTAIPDIETMRKIFPIDKILIGIVFCIAPPIASMPLOKNEFTVTD
IPFLIEPVMHTTAVFDFGVLCVAVGTVMITITISIGENE"
```

```
Alignment Scores:
Pred. No.: 861 Length: 290150
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x AP004825 (1-290150)
QY 1 ThrGlyGlyAsnSerGlySerProValphe 10
Db 118077 ACTGGTGTACTACGAGTTCCACCTGTATT 118048

RESULT 8
AP003132/C
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
genome, section 4/10.
ACCESSION
AP003132
VERSION
AP003132.2
KEYWORDS
GI:14349174
SOURCE
Staphylococcus aureus subsp. aureus N315 (sub-species:aureus N315,
strain:N315) DNA.
ORGANISM
Staphylococcus aureus subsp. aureus N315
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1
AUTHORS
Kuroda,M., Ohta,T., Uchlyama,I., Baba,T., Yuzawa,H., Kobayashi,I.,
Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M.,
Matsunaru,H., Maruyama,A., Murakami,H., Hosoyama,A.,
Mizutani-Oi,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C.,
Sekimizu,K., Hiraoka,H., Kihara,S., Goto,S., Yabuzaki,D.,
Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C.,
Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiratazu,K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
Lancet 357 (9264), 1225-1240 (2001)
JOURNAL
MEDLINE
21311952
PUBMED
11418146
REFERENCE
2 (bases 1 to 298050)
AUTHORS
Aoki,K., Oguchi,A., Hosoyama,A., Nagai,Y., Kuroda,M., Hiratsuki,K.
and Kikuchi,H.
Direct Submission
Submitted (30-JAN-2001) AKIO Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center, 2Chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nt.e.go.jp, URL:http://www.dio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
On Jun 12, 2001 this sequence version replaced gi:13700734.
COMMENT
FEATURES
location/Qualifiers
```

```
source
1. 298050
/organism="Staphylococcus aureus subsp. aureus N315"
/strain="N315"
/sub_species="aureus N315"
/db_xref="taxon:158879"
280. .432
/gene="SA0792"
280. .432
/gene="SA0792"
/note="ORFID:SA0792"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB42031.1"
/db_xref="GI:13700735"
/transl_table="MKSRSKQPPNKYVFAFKPYLLTLXLAIFITLLYLYTSGSDTHNN
FYNEF"
448. .1905
/gene="dltA"
448. .1905
/gene="dltA"
/note="ORFID:SA0793"
/codon_start=1
/transl_table=11
/product="D-alanine-D-alanyl carrier protein ligase"
/protein_id="BAB42032.1"
/db_xref="GI:13700736"
/transl_table="MTDIINKIQAFADANPOSIAVRHTTDELTYYOQIMDESKLAHRL
QSKRPMLIFGMSPYMIVGMIGAIKACGYVPVDSIPEDRIKMIINKVOEFENT
TDESFELEGEVETIEDIKTSQDPVIFPSQIKNDVTYTIETSGSGEPKGVQIVAS
LVOETFMMLINKSGNQOQMLNOAFPSFDLSVMAIYPCIASGGLNLVDKMKINRKL
LNEMLTAPININWSTPSEMEKCLPILNEQVGSINAEFFCGEILPHRAKALVSR
PSPATVNTYGTTEATVAVTSIOITQETIDQYPTLVGVERGARLSTDDCELYIEG
OSVSGIYKNDQKTRAEVNPDDGICRTYHTGDAKAKENQGMFTQGRIDQIKINGRME
LEIETQKQSEFEVFAIVPVYKNDKVIHLGALVPTTEVDNAEMTKINKDLKSR
LPEYMIIPKFEEMQLPLTSNGKIDRKRIAEVING"
1902. .3116
/gene="dltB"
1902. .3116
/gene="dltB"
/note="ORFID:SA0794"
/codon_start=1
/transl_table=11
/product="DltB membrane protein"
/protein_id="BAB42033.1"
/db_xref="GI:13700737"
/transl_table="WIPYGDFTFFLLIALILPVIILGFGKRSITYGVTAEMIVL
IFSSDKHMLFDQKYLVSQILSEFIIVVQVLLIMFYHRSKPKNNSKRVYVWASIL
PLAVKVLQSTWLGQIHFHESKLIIEVFGISYVTERKSVOLIMEIDGSIKEIKY
WLIQIFSEFPTISSPIDRKYRFYKDKKVPYTGNEYRELVLKAIHMLIGLYKYV
AVFITVAILMPLQDLHGPNVLMYVAYSLYLPFPDAGYSLEFAIFSYLIGIKTPPN
FDKPRKANKDFNNRHHMTLSFWRDCIYNSLIFYMSKRLIKSQPMSNAPLINF
FTGIMHGIEVYIYVIGLHALLTGYGIERMKRKHHPRMONGFTTALSTVITHEV
TFGLIFSGKLI"
3134. .3370
/gene="dltC"
3134. .3370
/gene="dltC"
/note="ORFID:SA0795"
/codon_start=1
/transl_table=11
/product="D-alanyl carrier protein"
/protein_id="BAB42034.1"
/db_xref="GI:13700738"
/transl_table="MEFRQVNLIAEVAENDIYKENDPVEIFEEGIIDSFOTVGLL
EIQNKLDIEVSLIMDFRDQEWATPNKIVALEELR"
3367. .4542
/gene="dltD"
3367. .4542
/gene="dltD"
/note="(glycerophosphate chain)
ORFID:SA0796"
/codon_start=1

gene
/transl_table=11
/product="poly D-alanine transfer protein"
/protein_id="BAB42035.1"
/db_xref="GI:13700739"
/transl_table="WKLKPELPILISGAVETVLELLPASPWFTGLVNEKTVEDNRTSLT
DOVLKQGLIQQKLYESNKYPIYSGSSELGKDDPPADALAKHNKKAFLLAGAGST
DLINAVALEAQDLKGRKLTETIISPOWFTNHGLINQNDAMSQTOINQMFQKNS
TEIKRVAORLIQPPHVNKEIYKSAKNPKETKDSYISGFERQOLIKTEAIKSLFMS
DKSPLEHAKPATKPDASWDEMKOKAVEYKADTTSNKRGIQDQYVILITQESKKRYRD
YEFNVSPEFQDDELIVKTRAGADVOYVISPNSGVYDHIIGIDKERQAVYKRIHS
TVVDNGKLYDMTDKDYERYVYISDAVNHGKGMVYMDQIAKHMKGEPPVDPKPN"
complement(4805. .5047)
/gene="SA0797"
complement(4805. .5047)
/gene="SA0797"
/note="ORFID:SA0797"
hypothetical protein, similar to nitrogen fixation protein
NifU"
/codon_start=1
/transl_table=11
/protein_id="BAB42036.1"
/db_xref="GI:13700740"
/transl_table="MPTEDTTFDQVAEIERLRPFLRDGDCSLIDVEDGIYKQL
HGACGTCPSSTITTKAGIERALHEEVPIVEQVF"
5163. .5471
/gene="SA0798"
5163. .5471
/gene="SA0798"
/note="ORFID:SA0798"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42037.1"
/db_xref="GI:13700741"
/transl_table="WVYVGADYTCASVNAPTSNDYDMLQPLKRRKPNISFKYTYI
DITKDNMLTDHDLQFIERIEDLEFLYLTMDNEYVADGYIQTKQIRFDQKLVNE"
complement(5530. .6594)
/gene="SA0799"
complement(5530. .6594)
/gene="SA0799"
/note="ORFID:SA0799"
hypothetical protein, similar to NADH dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB42038.1"
/db_xref="GI:13700742"
/transl_table="MKNVLILGGYGNMRINSRLTTSLPDYTVTLVDRMPFHGLK
PEYALAACTSKDQVYRPNHPQVNTYVGEINDLDAQIVSNGSKLDYDELIIGL
GCEDYVHVPQAEETHYSTIONLSKARDPFHSISELPBAGAGIVAGISGIEASELR
ESRSLEIYVLDROGRILIRNPPEKISKYAKKFAKNNTYVPSNINRNVBGRKITNCD
EPKIDILVWTRAGIOPVEVVRNPLIDINSNGVAVNOHYOFTYRNATVVGACALPH
APSQAIAVQGGQIADVLKQWLNPELPDKMEELKVOGIVGSLGDKGFAYIMDRVT
GRLASILKSGVIMLYKHNG"
6911. .7147
/gene="SA0800"
6911. .7147
/gene="SA0800"
/note="ORFID:SA0800"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42039.1"
/db_xref="GI:13700743"
/transl_table="MNPIVERCLSMARGGDVFNQLENPDVDVLEXGCLTHGICS
AGLYALVNGDIYEGDSPPELLQNIYAHIKETWIF"
7160. .7519
/gene="SA0801"
7160. .7519
/gene="SA0801"
/note="ORFID:SA0801"
/codon_start=1
/transl_table=11
```

/product="conserved hypothetical protein"
/protein_id="BAB42040.1"
/db_xref="gi:13700744"
/translation="MTVLTLEAAAYEVKDMKANEMPDGYLIKIVNGGGCGTLYGM
GAEEAPGENDVLETFGLKVLVDKEDAPVLANGTITDFKOSLMGGGFQIDNPALASCG
CGSSFRKAVKAGNPENC"
7973. .9181
/gene="SA0802"
7973. .9181
/gene="SA0802"
/note="ORFID:SA0802
hypothetical protein, similar to NADH dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB42041.1"
/db_xref="gi:13700745"
/translation="MAODRKVLVAGYAGAGTQYTKLQKASTREAITLINKNEYH
YEATVLEASAGTLEEDVLEFVESVLKDKVNEVQAEVTKIDRDARKVETQGLYDF
DILVALFVSETFEGIEGMDHAFQIENVITARELSRIIEKRFANYAASKEDNDLS
ILVGGAGFTGVEFLGELTDRIPELCSKYGVQDNKVKITCVEAAPKMLPMFSEELVNAH

Alignment Scores:
Pred. No.: 878 Length: 298050
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x AP003132 (1-298050)
Oy 1 ThrGlyGlyAsnSerGlySerProValphe 10
Db 123979 ACTGGTGTAACTACAGTTCACCTGATTTT 123950

RESULT 9
AP003361/c 347235 bp DNA linear BCT 07-FEB-2002
LOCUS Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
DEFINITION sequence, section 4/9.
ACCESSION AP003361 BA000017
VERSION AP003361.2 GI:14246761
KEYWORDS
SOURCE Staphylococcus aureus subsp. aureus Mu50 (sub-species:aureus Mu50,
strain:Mu50) DNA.
ORGANISM Staphylococcus aureus subsp. aureus Mu50
Bacteria; Firmicutes; Bacillales; Staphylococcus.

REFERENCE
AUTHORS 1
Kuroda, M., Ohta, T., Uchiyama, T., Baba, T., Yuzawa, H., Kobayashi, I.,
Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Lian, J., Ito, T., Kanamori, M.,
Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,
Mizutani, O., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C.,
Sekimizu, K., Hirakawa, H., Kohara, S., Goto, S., Yabuzaki, J.,
Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,
Shiba, T., Hatori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus

TITLE
JOURNAL Lancet 357 (9264), 1225-1240 (2001)
MEDLINE 21311952
REFERENCE 2 (bases 1 to 347235)
AUTHORS Ohta, T.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology, 1-1-1 Ten-no-dai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail: tohtasakura.cc.tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)
On May 29, 2001 this sequence version replaced gi:13875305.

COMMENT
FEATURES Location/Qualifiers
1..347235
/organism="Staphylococcus aureus subsp. aureus Mu50"
/strain="Mu50"
/sub-species="aureus Mu50"

/db_xref="taxon:158878"
35. .1021
/gene="SAV0992"
35. .1021
/gene="SAV0992"
/codon_start=1
/transl_table=11
/product="Oligopeptide ABC transporter ATP-binding protein
homologue"
/protein_id="BAB57154.1"
/db_xref="gi:14246762"
/translation="MNNVLLEVKDLETSLKINNEMLATVENISELSGEVLIVGES
CGGSIISKSIKILPEKISKLSGEVIFDGRKIDITLNEKQIDIRGNDIMIFQEPM
TALNPVETIKNOLVESIKSHKISKKEANNAKDLKLVGIARODEILNSVPHOLSG
MRORVIMATNISCSPKILLADEPTALDVTIOAOTIDLEKQETQMAIMTTHDS
VVAERCDKVLVYNAQIVYEREGIKETILNPPHPTTOKLSTIPKLEQKLEITIGI
VPSIOAFHNKCRFRANCKRKLDCNNOSPMMHVCEDVIVCHLYKNYKFI"
1024. .2004
/gene="apbf"
1024. .2004
/gene="apbf"
/note="SAV0993"
/codon_start=1
/transl_table=11
/product="Oligopeptide transport system ATP-binding
protein Apbf homologue"
/protein_id="BAB57155.1"
/db_xref="gi:14246763"
/translation="MENTLEVNQIKKIKITGGLQTOYKAVDNDVSFSIKKQTFG
LVESGCGSTLGVKIVIRLEDATSGSIIVNEDITRLQKLRKSRQOYQVIFDDPA
SLNPMOWAGDIISPEILNPKLPKEIKKENVLYLKCYLSEDAVYKKAHFGSGORQ
RVGIARALALRPSLIVADEPVSALDVOSOVNLKLOBOFKLSYLFIAHDSVVK
HISDVGMYLGHIVETIASDKETIENPMPHTKALISSIPQIDKHNNRITLKKEPLS
PSNPQSGPFRTPCIADKCKENIPQLKDIGNEHQVACFYVDKVDLNA"
1997. .2959
/gene="oppb"
1997. .2959
/gene="oppb"
/note="SAV0994"
/codon_start=1
/transl_table=11
/product="probable oligopeptide transport system permease
protein"
/protein_id="BAB57156.1"
/db_xref="gi:14246764"
/translation="WKLILKRLGLMPLILLISIVPSLAITPGDPFSDLOGKIK
QEAINAOREKGLNDSIPHOYIRWNVNHTHDLDESIRKRPVADIVIEERIPNTLLG
AMSLITYIISFLAIGITSGRYSYSLDTVQIFVYLMALPISFAGVFAIFPSFEIG
WEPFGSDVIMLKEGTEFYMSKIYHFFLPALITGLSLSTAGYIOYLRNDITENSKDY
VLTFRSKGLSMNKIYLNKHIILRNSLIPITTFGADVISILGAVITETIIFYNGIGKLF
LESYIGDQYPLMMLTLFFSFLGLGNLISDIYTGFIIDPRIRSN"
2971. .3852
/gene="SAV0995"
2971. .3852
/gene="SAV0995"
/note="similar to oligopeptide transport system permease
protein Oppc"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57157.1"
/db_xref="gi:14246765"
/translation="MONKSKSPKIALSRFTHNKIMASTIFLLITIVSIAPLAP
FPVNOODLNTKSGMTNONTIGTSGGRDNFSRLLYAGRISLSIGITSTMGLIGIT
VGVISGFRGGIVDTLMLKITEFVLPFLPFAIVLALGKINPVSATITVLVIT
VLSWGLARLVGRVQLQENEFVLAASKISTPYKIIILKHLINLSIVYQATLFL
AGMIVVSGLSFLFGISKAIPSGNMMLSDAEOGDVISGKPMWIMPPAIMTITLISI
NFVGEGLKDAFNPGRR"
complement(3894. .4883)
/gene="trps"
complement(3894. .4883)
/note="SAV0996"

gene
CDS

gene
CDS

gene
CDS

gene
CDS

```
/codon_start=1
/transl_table=11
/product="tryptophanyl-tRNA synthetase"
/protein_id="BAB57158.1"
/db_xref="GI:14246766"
/translation="METLEFSGIOPSGIPTIGNYIGALKOFVQNDYQDCYCIYDOHA
ITMPDRLKRTKQROLAAYIYASIGIDPKATLFIQSEPAHVGAGMLTIYASGL
ERMVOYKDKAKQAVGIPAGLIYATPLMAADIVYNTNIVPGDOKROHETLRNLVD
RNSRYNDVLYKPEIRMKVGRVNSLDDPIRKMSKSDNAKNFISLDERVAKKI
KSAVYDSGIIKEDRNKPGITNLISITAGLDMPFKDIEAKYEGSGYGRKGDIAEI
VKAFLVEQEKYESSEYNSDKLDLIDQGRNAHKSFYKVRKMEKMGIGRRK"
/gene="SAV0997"
/5177..5572
/gene="SAV0997"
/5177..5572
/gene="SAV0997"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB57159.1"
/db_xref="GI:14246767"
/translation="MVTLPFSPTSCSKRANLQEHPIPTERNIESEHITIDRIKO
IKMTEDGDEIISTRSKYQKLVNDIDSLPLQDLYSLIDNPGLLRPIILDNRLO
VGNDEIRRLPRKVRTEOLOAQRMD"
/5943..6662
/gene="SAV0998"
/5943..6662
/gene="SAV0998"
/5943..6662
/gene="SAV0998"
/5943..6662
/feature="similar to negative regulator of genetic competence
Meca"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57160.1"
/db_xref="GI:14246768"
/translation="WRIERVDITVYKLEITFYSDIEAKGFSREDLMTNKRGEFPWMS
MADEINEEDFVEGELMIOVAHFEKGVAVITISKNEEDMNSDDADQDEQVOEL
LAQULEGDOLEELFEORTKEKQAGSRSKAKNRRTIIVKNDIEDYINVAYS
NPITFEFDLLYMGVGYVAVHPDSHDQEVINDYSQLEFAFPTPTREYVLYNDYA
KTIKSHNTAQRVRRFPETTE"
/6783..7769
/gene="SAV0999"
/6783..7769
/gene="SAV0999"
/6783..7769
/feature="similar to transcription factor"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB57161.1"
/db_xref="GI:14246769"
/translation="MLVALNEKEKERVLTATLRKTOYFCPVQCGKOVILKRGIVSHF
AHKHLABOKCFNNETIKHKSKLLIAOMIOOQGCVELEPELKEIKQIPDILINKYV
IELQSPILPYKQIILORTGLKRMGTKVSMLNDVDYCHNKVFNHFSLFNPTIRKL
HTFNLEKQOIMFQOIQYLGSHKRYAEKRNKIIELFEAPDQYAAVYKLSFALNOY
IKYCRWQNSVLEPTLSAMYQOLTDQEVYVNGYLFEPQIYIENPIEMQLOVDMRK
NGSKLVNDNINYEFLKKFIVALESKTALIERKLNNYNINICSDRNDVOJLF"
/7817..9625
/gene="SAV1000"
/7817..9625
/gene="SAV1000"
/7817..9625
/codon_start=1
/transl_table=11
/product="thimet oligopeptidase homologue"
/protein_id="BAB57162.1"
/db_xref="GI:14246770"
/translation="WSQQLSRBEQRRKYPEYTWDLITTFKQDEAEAFKYEVENELKQ
EEOFGHIGDSAEETLYNALLEDLTGLTKLEYVYVAHLKODDITNDKYTGESRAHQ
LIIRKSAMSFLEPEILQIDEDKIQSFVNSYDKLOKPAFDLKLINERPHIADAEK
LITEADALSTPSNYYGMEFNADIVFEADIDKGNAPHLTQGTFLIKYLESDDRKLRES
AFRNYKAYGANNNTIGATLACGCKKNPNARTHYKTKAREKALSNNHNPENYVNVY
KTVKTYLPILHRYTELKREKLGDLDDLMKTDITLTKDKIKEMPEYEAKEMWKLKLE
MGEETLVNVAKEGLNNRWVDYENKGRSGYSSGHLNPNPILLMWSNTISDLTYLVE
FEGSHASYFSRKPQSPNSDVTTFVAVASVCNALLSDYMDKHLDEKRLILLNOE
LERFRATLFRQTMFAEFKHIHAIIEAGEPIPTPTMNEVAKLNKLYFGDSVETDEDI
```

```
gene
CDS
/translation="SKEMSRIPHFMYNYYOYATGYSAAOSLSHQLITECKPAVDRIINEFLKKGSSNYP
ILKMGADMTTPPEIPEACEVFEKCKLNAFEKIMKA"
/complement(10085..10639)
/gene="yjbM"
/complement(10085..10639)
/gene="yjbM"
/feature="SAV1001"
/codon_start=1
/transl_table=11
/product="Gnp pyrophosphokinase"
/protein_id="BAB57163.1"
/db_xref="GI:14246771"
/translation="MKOLSVRSKTEKIYVTCVNEIIPKRDITTESMIDQIONAG
IDLEVFKRDLOKSKTESLKDIDHARMEIQQAPSLVFESEDDHEEGLKYGALPYH
IYITITNELMGKPILEKNLPPKUEYTIQDQQLVWEELITTEWPEKLLNKLKLAIO
QKIEKLKYPDGDFFWKSMPKIKSK"
/complement(10575..10901)
/gene="SAV1002"

Alignment Scores:
Pred. No.: 982 Length: 347235
Score: 54.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x AP003361 (1-347235)
OY 1 ThrGlyGlyAsnSerGlySerProValPhe 10
Db 56369 ACTGGTGTAAGTCAAGTTCACCTGATATT 56340

RESULT 10
SEP305145 657 bp DNA linear BCT 25-JAN-2002
LOCUS Staphylococcus epidermidis partial esp gene for extracellular
DEFINITION
serine proteinase.
ACCESSION
AJ305145
VERSION
AJ305145.1 GI:12580903
KEYWORDS
esp gene; extracellular serine proteinase.
SOURCE
Staphylococcus epidermidis.
ORGANISM
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1 Dubin,G., Chmiele,D., Mak,P., Rakwalaska,M., Rzychon,M. and Dubin,A.
Molecular cloning and biochemical characterisation of proteases
from Staphylococcus epidermidis
Biol. Chem. 382 (11), 1575-1582 (2001)
JOURNAL
MEDLINE
21618051
PUBMED
11767947
REFERENCE
2 (bases 1 to 657)
Dubin,G.
Direct Submision
Submitted (19-JAN-2001) Dubin G., Institute of Molecular Biology,
Jagiellonian University, ul. Mickiewicza 3, Krakow 31-120, POLAND
FEATURES
source
1. 657
/organism="Staphylococcus epidermidis"
/strain="6746"
/db_xref="taxon:1282"
1..654
/gene="esp"
<1..654
/gene="esp"
/codon_start=1
/translation="
/feature="experimental
/translation="
/feature="extracellular serine proteinase"
/protein_id="CAC27157.1"
/db_xref="GI:12580904"
/db_xref="SPTREMBL:Q9AJX0"
/translation="SVILPNNRRHQIFHTQGHDAVSFIYPIIDGYSGSGVYVGE
```

NEILTKNVNCAKGNPRNINISVHPSAKNENDYPNGKEVGOETIIPYNGSLAILFVSP
NEHNOHQVVKPATISSNTDRIINENTVTGYPEDKPLATMSESVGVVYIGGEELR
YDLSTVGNGSGSPVNGKNQVYIGHYGVGNKYNSSVYINDVQOFLNNIPDINIQ"

sig_peptide
4.3
/gene="esp"
4.651
/gene="esp"
/product="extracellular serine proteinase"
/evidence=experimental

mat_peptide
88 c 122 g 204 t

BASE COUNT 243 a 88 c 122 g 204 t
ORIGIN

Alignment Scores:
Pred. No.: 59.4 Length: 657
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x SEP305145 (1-657)

QY 2 GYGLYASnseryglyserprovalphe 10
|||||
Db 499 GGTGGAACCTCTGATCTCCAGTATT 525

RESULT 11
AX141641 849 bp DNA linear PAT 31-MAY-2001
LOCUS
DEFINITION Sequence 363 from Patent W00134809.
ACCESSION AX141641
VERSION AX141641.1 GI:14281692
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 849)
AUTHORS Kimmerly,W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 363 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
source location/Qualifiers
1. 849
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic nucleic acid sequence"

BASE COUNT 325 a 122 c 147 g 255 t
ORIGIN

Alignment Scores:
Pred. No.: 71.7 Length: 849
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 6 Gaps: 0

US-10-008-355-25 (1-10) x AX141641 (1-849)

QY 2 GYGLYASnseryglyserprovalphe 10
|||||
Db 694 GGTGGAACCTCTGATCTCCAGTATT 720

RESULT 12
AF269652/c 3189 bp DNA linear ECT 01-AUG-2000
LOCUS
DEFINITION Staphylococcus epidermidis strain SRI clone step.1015e12 genomic
sequence.
ACCESSION AF269652
VERSION AF269652.1 GI:9623548
KEYWORDS
SOURCE Staphylococcus epidermidis.
ORGANISM Staphylococcus epidermidis

Bacteria: Firmicutes: Bacillales: Staphylococcus.
1 (bases 1 to 3189)
REFERENCE Kimmerly,W.J., Taylor,J.David., Nelsen,A.J., Godlevski,M.M.,
AUTHORS Rubino,M.A., Nelson,F.J., Rivers,P.R., Torturella-Miller,I.,
Listendee,S., Ashanti,C., Alshuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis
genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3189)
AUTHORS Taylor,J.David., Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Torturella-Miller,I.,
Listendee,S., Ashanti,C., Alshuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Direct Submission
JOURNAL Submitted (22-May-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
source location/Qualifiers
1. 3189
/organism="Staphylococcus epidermidis"
/strain="SRI"
/db_xref="taxon:1282"
/clone="step.1015e12"

BASE COUNT 1081 a 596 c 502 g 1010 t
ORIGIN

Alignment Scores:
Pred. No.: 189 Length: 3189
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x AF269652 (1-3189)

QY 2 GYGLYASnseryglyserprovalphe 10
|||||
Db 1147 GGTGGAACCTCTGATCTCCAGTATT 1121

RESULT 13
AX144972/c 3189 bp DNA linear PAT 31-MAY-2001
LOCUS
DEFINITION Sequence 3694 from Patent W00134809.
ACCESSION AX144972
VERSION AX144972.1 GI:14283537
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 3189)
AUTHORS Kimmerly,W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 3694 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
source location/Qualifiers
1. 3189
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic nucleic acid sequence"

BASE COUNT 1081 a 596 c 502 g 1010 t
ORIGIN

Alignment Scores:
Pred. No.: 189 Length: 3189
Score: 49.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.74% Indels: 0
DB: 6 Gaps: 0

US-10-008-355-25 (1-10) x AX144972 (1-3189)			
OY	2	GlyGlyAsnSerGlySerProValPhe	10
Db	1147	GGTGAACCTCGATCCAGTATT	1121
RESULT 14			
AE004008/c	10689	bp	DNA linear BCT 15-JUN-2001
LOCUS	Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.		
DEFINITION	Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.		
ACCESSION	AE004008 AE003849		
VERSION	AE004008.1 GI:9106961		
KEYWORDS			
SOURCE	Xylella fastidiosa 9a5c.		
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xylella.		
REFERENCE	1 (bases 1 to 10689)		
AUTHORS	Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M., Araya,J.E., Bata,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facinani,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., Ho,P.L., Hohnselt,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P. and Marino,C.L.		
TITLE	The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis		
JOURNAL	Nature 406 (6792), 151-157 (2000)		
MEDLINE	20365717		
PUBMED	10910347		
REFERENCE	2 (bases 1 to 10689)		
AUTHORS	Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Bata,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facinani,A.P., Ferreira,A.J.S., Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S., Gomes,S.L., Gruber,A., Ho,P.L., Hohnselt,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M., Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Weidmanis,J. and Setubal,J.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil		
FEATURES			
source	1. 10689 location/Qualifiers /organism="Xylella fastidiosa 9a5c" /db_xref="taxon:160492" /clone="9a5c" complement(119..367)		
gene			

CDS	gene	complement(119..367) /gene="XF1871" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84677.1" /db_xref="GI:9106962" /translation="MTMTLKEDKTEIIDEORIQIAKLSETAKIQAEIHVNAHTQKL MKETLKPTESKWPVVGSSSLIAGATATLFIKLFH" complement(430..618) /gene="XF1872" complement(430..618) /gene="XF1872" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84678.1" /db_xref="GI:9106963" /translation="MMEBARATAPQGLAASDPQRTDAPKKTGKRKRYPPTGSASI IYLRHVLRQQRKASP" 941..1288 /gene="XF1873" 941..1288 /gene="XF1873" /note="similar to GI1479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %): identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAF84679.1" /db_xref="GI:9106964" /translation="MSKKIRPQITGTGMAAGPQFINGKLYMPERTIEMOLRWLST CTLAREQKMEBARATAPQGLAASDPQRTDAPKKTGKRKRHANEASTIOLRO VLIQRQQRKASP" 1351..1662 /gene="XF1874" 1351..1662 /gene="XF1874" /note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAF84680.1" /db_xref="GI:9106965" /translation="MTMTLKEDKTEIIDEOKIQIAKLISEAKIQAEIHVNTQEM NAHNIEMNARDEMNARTQKLEEFQKSKESQWPVVGSLAAGATATLFIKLF H" complement(1839..2918) /gene="XF1875" complement(1839..2918) /gene="XF1875" /note="similar to SPIP03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %): identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder" /codon_start=1 /transl_table=11 /product="phase-related protein" /protein_id="AAF84681.1" /db_xref="GI:9106966" /translation="MLNVTGAPCGKTLVAVDMILROIETDKSLVKSQAVPRRYTD IEGFVAVRRLTGTVVOSAPEDWRTTQGSIVTIDEAHKMPAGRPERSDDPRVCDL DTHRGGVGLMEVQTQMPKIHHLRRLVGEHVLNRAMGLQTAGLYRMSRQDDPYDI HOREKAEVEVMPKDRYALVASSSTLHVSHFRIPKRYWASLSVCVCSVIGLVFWM YSPAHLSAASSVAGSQASLRAAPSLSSRSLVSGMRTYAVLLETESAVPTLSGCV
-----	------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

```

SSDRSCRCFNTDGYOIMDSVVECRRLASPLPFNNYHAYVTSSSSAAAVSSSPSSSPSS
SSSVLSSSSVAPILGSSVSPSSH"
gene      complement(2918..3202)
CDS       /gene="XF1876"
          /complement(2918..3202)
          /note="similar to SP|P03625 (percent identity: 36 %/query
alignment coverage: 85.1 %/subject alignment coverage:
86.0 %): identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="phage-related protein"
          /protein_id="AAF84682.1"
          /db_xref="GI:9106967"
          /translation="MSWLAMLRNIIIGOTVARGIVAGIALVTTPLIPIVTSALNL
IYSKSGISADVLNIALMLGREGALSTIGSAMLTRALQSLHVGIVKATT"
gene      complement(3195..4412)
CDS       /gene="XF1877"
          /complement(3195..4412)
          /note="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAF84683.1"
          /db_xref="GI:9106968"
          /translation="MCSRSFLYFISAGLTIIVISPFCLPQNVPVVEVPTVYR
SSVSDGVSTATFEARSTALVNGRVYITVPDISASTIGSLAKVAVRGMAFVNVYS
MISGLINGAGVYIDELTHEVMSGPAKLEIPVTVAMWQRPDGHVFYSVPQGLIAP
INAYOSLSLPPQIVLTSSGSPSSSERMLYHLEGGEYIGHLSKTDQSVPDGSLAP
SVILDTDLGOLVRSDPSTVNAVLTDSOTGAVLLPEIYSALPKLRSLDEDELYSKAP
DOQPSGGASPPSSPTGAWPSCFIMQSCVAFIDMKSDEFLKPLVPPDPVYVDKL
PQAKTMSGIGBACPSPTTIIIEFSGYKTSVEISGYPCDFCALMRVYIATILTA
AYIACFSGKVNW"
gene      complement(4481..4702)
CDS       /gene="XF1878"
          /complement(4481..4702)
          /note="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAF84684.1"
          /db_xref="GI:9106969"
          /translation="MKSITDRGIAFSLSTSLVFADAGASFDACAATVATCAIS
GAVALIGAKKLPAIISVGWMLKGAIRG"
gene      complement(4724..4924)
CDS       /gene="XF1879"
          /complement(4724..4924)
          /note="hypothetical protein: identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAF84685.1"
          /db_xref="GI:9106970"
          /translation="MAQQLMSLYCTQYVEANTSQQAQMMVPPSLPISTEDVRILL
PHIVMCFIVAWGFHLFTYVRD"
gene      complement(4931..5236)

```

```

Alignment Scores:
Pred. No.:      656      Length:      10689
Score:          48.00     Matches:      9
Percent Similarity: 100.00%  Conservaive: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match:      88.89%     Indels:      0
DB:              1         Gaps:      0
US-10-008-355-25 (1-10) x AE004008 (1-10689)

```

```

Oy      1  ThrGlyGlyAsnSerGlySerProval 9
Db      8561 ACTGGCGGCACTCCGATCGCCAGTCC 8535

RESULT 15
AC121125
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 66762)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
Anderson,S., Barrera,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Illiev,I., Johnson,R., Jones,C.,
Kamel,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mieng,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Rella,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testafaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,M., Young,G.,
Zainoun,J., Zemke,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (15-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
ALL repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

```

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I23222
Center clone name: 10_F_17
-----
* NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 717: contig of 717 bp in length
* 718 817: gap of 100 bp
* 818 1529: contig of 712 bp in length
* 1530 1629: gap of 100 bp

```

* 1630 2339: contig of 710 bp in length
* 2340 2439: gap of 100 bp
* 2440 3146: contig of 707 bp in length
* 3147 3246: gap of 100 bp
* 3247 3965: contig of 719 bp in length
* 3966 4065: gap of 100 bp
* 4066 4793: contig of 728 bp in length
* 4794 4893: gap of 100 bp
* 4894 5605: contig of 712 bp in length
* 5606 5705: gap of 100 bp
* 5706 6422: contig of 717 bp in length
* 6423 6522: gap of 100 bp
* 6523 7192: contig of 670 bp in length
* 7193 7292: gap of 100 bp
* 7293 8001: contig of 709 bp in length
* 8002 8101: gap of 100 bp
* 8102 8806: contig of 705 bp in length
* 8807 8906: gap of 100 bp
* 8907 9620: contig of 714 bp in length
* 9621 9720: gap of 100 bp
* 9721 10406: contig of 686 bp in length
* 10407 10506: gap of 100 bp
* 10507 11232: contig of 726 bp in length
* 11233 11332: gap of 100 bp
* 11333 12055: contig of 723 bp in length
* 12056 12155: gap of 100 bp
* 12156 12887: contig of 732 bp in length
* 12888 12987: gap of 100 bp
* 12989 13716: contig of 729 bp in length
* 13717 13816: gap of 100 bp
* 13817 14539: contig of 723 bp in length
* 14540 14639: gap of 100 bp
* 14640 15344: contig of 705 bp in length
* 15345 15444: gap of 100 bp
* 15445 16157: contig of 713 bp in length
* 16158 16257: gap of 100 bp
* 16258 16968: contig of 711 bp in length
* 16969 17068: gap of 100 bp
* 17069 17781: contig of 713 bp in length
* 17782 17881: gap of 100 bp
* 17882 18602: contig of 721 bp in length
* 18603 18702: gap of 100 bp
* 18703 19399: contig of 697 bp in length
* 19400 19499: gap of 100 bp
* 19500 20231: contig of 732 bp in length
* 20232 20331: gap of 100 bp
* 20332 21036: contig of 705 bp in length
* 21037 21136: gap of 100 bp
* 21137 21862: contig of 726 bp in length
* 21863 21962: gap of 100 bp
* 21963 22701: contig of 739 bp in length
* 22702 22801: gap of 100 bp
* 22802 23513: contig of 712 bp in length
* 23514 23613: gap of 100 bp
* 23614 24326: contig of 713 bp in length
* 24327 24426: gap of 100 bp
* 24427 25144: contig of 718 bp in length
* 25145 25244: gap of 100 bp
* 25245 25960: contig of 716 bp in length
* 25961 26060: gap of 100 bp
* 26061 26778: contig of 718 bp in length
* 26779 26878: gap of 100 bp
* 26879 27574: contig of 696 bp in length
* 27575 27674: gap of 100 bp
* 27675 28396: contig of 722 bp in length
* 28397 28496: gap of 100 bp
* 28497 29173: contig of 677 bp in length
* 29174 29273: gap of 100 bp
* 29274 30002: contig of 729 bp in length
* 30003 30102: gap of 100 bp
* 30103 30825: contig of 723 bp in length
* 30826 30925: gap of 100 bp
* 30926 31647: contig of 722 bp in length

* 31648 31747: gap of 100 bp
* 31748 32459: contig of 712 bp in length
* 32460 32559: gap of 100 bp
* 32560 33277: contig of 718 bp in length
* 33278 33377: gap of 100 bp
* 33378 34086: contig of 709 bp in length
* 34087 34186: gap of 100 bp
* 34187 34902: contig of 716 bp in length
* 34903 35002: gap of 100 bp
* 35003 35714: contig of 712 bp in length
* 35715 35814: gap of 100 bp
* 35815 36543: contig of 729 bp in length
* 36544 36643: gap of 100 bp
* 36644 37367: contig of 724 bp in length
* 37368 37467: gap of 100 bp
* 37468 38191: contig of 724 bp in length
* 38192 38291: gap of 100 bp
* 38292 39012: contig of 721 bp in length
* 39013 39112: gap of 100 bp
* 39113 39829: contig of 717 bp in length
* 39830 39929: gap of 100 bp
* 39930 40639: contig of 710 bp in length
* 40640 40739: gap of 100 bp
* 40740 41453: contig of 714 bp in length
* 41454 41553: gap of 100 bp
* 41554 42275: contig of 722 bp in length
* 42276 42375: gap of 100 bp
* 42376 43090: contig of 715 bp in length
* 43091 43190: gap of 100 bp
* 43191 43905: contig of 715 bp in length
* 43906 44005: gap of 100 bp
* 44006 44688: contig of 683 bp in length
* 44689 44788: gap of 100 bp
* 44789 45493: contig of 705 bp in length
* 45494 45593: gap of 100 bp
* 45594 46316: contig of 723 bp in length
* 46317 46416: gap of 100 bp
* 46417 47136: contig of 720 bp in length
* 47137 47236: gap of 100 bp
* 47237 47927: contig of 691 bp in length
* 47928 48027: gap of 100 bp
* 48028 48749: contig of 722 bp in length
* 48750 48849: gap of 100 bp
* 48850 49563: contig of 714 bp in length
* 49564 49663: gap of 100 bp
* 49664 50375: contig of 712 bp in length
* 50376 50475: gap of 100 bp
* 50476 51193: contig of 718 bp in length
* 51194 51293: gap of 100 bp
* 51294 52013: contig of 720 bp in length
* 52014 52113: gap of 100 bp
* 52114 52887: contig of 774 bp in length
* 52888 52987: gap of 100 bp
* 52988 53717: contig of 730 bp in length
* 53718 53817: gap of 100 bp
* 53818 54547: contig of 730 bp in length
* 54548 54647: gap of 100 bp
* 54648 55374: contig of 727 bp in length
* 55375 55474: gap of 100 bp
* 55475 56189: contig of 715 bp in length
* 56190 56289: gap of 100 bp

Alignment Scores:

Pred. No.: 7.34e+03 Length: 66762
Score: 45.00 Matches: 8
Percent Similarity: 90.008 Conservative: 1
Best Local Similarity: 80.008 Mismatches: 1
Query Match: 83.33% Indels: 0
DB: 2 Gaps: 0

US-10-008-355-25 (1-10) x AC1211125 (1-66762)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10

||||| |||||||||||||
Db 57225 ACAGGAATTAACAGCGCTCTCCCTATT 57254

Search completed: December 20, 2002, 15:01:46
Job time : 1615.74 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:10:29 ; Search time 38.4211 Seconds

(without alignments)
79.820 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TCGNSGSPVF 10

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 segs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seg length: 0

Maximum DB seg length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODE=framer.p2n.model -DEV=xlh
-Q=/cgn2_1/uspto.spool/US10008355/rnuc.17122002.112337.14581/app.query.fasta_1.398
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human4.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPM=p2n -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US10008355.ecgn_1.17-sequnal.17122002.112337.14581 -MCPD=6 -ICPD=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued_Patents_NA.*

1: /cgn2_6/prodata/1/ina/5A.COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B.COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A.COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B.COMB.seq.*
5: /cgn2_6/prodata/1/ina/PCTUS.COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	1974	4	US-09-321-017B-726
2	49	90.7	936	4	US-09-134-001C-892
3	48	88.9	2384	4	US-09-221-017B-1045
4	44	81.5	758	4	US-09-071-035-427
5	44	81.5	888	4	US-09-071-035-425
6	43	79.6	1448	1	US-08-035-634-1
7	40	74.1	743	4	US-08-998-416-856
8	39	72.2	240	3	US-08-626-169-2
9	39	72.2	240	3	US-09-164-307-2
10	39	72.2	2774	1	US-07-723-002C-5
11	38	70.4	531	4	US-09-222-575-123
12	38	70.4	663	1	US-07-624-113-3

13	38	70.4	3283	4	US-09-061-709-8	Sequence 8, Appl
14	38	70.4	3412	4	US-09-061-709-6	Sequence 6, Appl
15	37	68.5	39	1	US-08-225-224-31	Sequence 31, Appl
16	37	68.5	39	5	US-08-722-258-31	Sequence 31, Appl
17	37	68.5	39	5	PCT-US95-04468-31	Sequence 31, Appl
18	37	68.5	42	1	US-08-225-224-28	Sequence 28, Appl
19	37	68.5	42	3	US-08-722-258-28	Sequence 28, Appl
20	37	68.5	42	5	PCT-US95-04468-28	Sequence 28, Appl
21	37	68.5	633	1	US-08-664-596B-21	Sequence 21, Appl
22	37	68.5	3663	4	US-09-499-884-11	Sequence 11, Appl
23	36	66.7	184	2	US-08-887-365-32	Sequence 32, Appl
24	36	66.7	454	2	US-08-680-326-118	Sequence 118, App
25	36	66.7	686	4	US-09-221-017B-38	Sequence 38, App
26	36	66.7	1291	2	US-08-887-365-35	Sequence 35, Appl
27	36	66.7	3958	2	US-07-952-853-21	Sequence 21, Appl
28	36	66.7	3958	2	US-08-914-848-21	Sequence 21, Appl
29	36	66.7	5045	4	US-09-390-721-1	Sequence 1, Appl
30	36	66.7	5045	4	US-09-390-721-3	Sequence 3, Appl
31	36	66.7	5399	1	US-08-064-121-1	Sequence 1, Appl
32	36	66.7	5399	1	US-08-478-015-1	Sequence 1, Appl
33	36	66.7	5399	3	US-08-475-975-1	Sequence 1, Appl
34	36	66.7	5399	3	US-09-084-889-1	Sequence 1, Appl
35	36	66.7	6555	1	US-08-351-413-2	Sequence 1, Appl
36	36	66.7	6555	2	US-09-025-583-2	Sequence 2, Appl
37	36	66.7	6854	4	US-09-194-905-7	Sequence 7, Appl
38	36	66.7	7566	2	US-08-232-016-23	Sequence 23, Appl
39	36	66.7	7639	2	US-08-232-016-22	Sequence 22, Appl
40	36	66.7	12687	1	US-08-676-169-1	Sequence 1, Appl
41	36	66.7	12687	3	US-08-981-459-1	Sequence 1, Appl
42	36	66.7	12687	4	US-09-063-431A-1	Sequence 1, Appl
43	36	66.7	15397	2	US-08-673-768-1	Sequence 1, Appl
44	36	66.7	15397	2	US-08-673-768-1	Sequence 1, Appl
45	36	66.7	24595	6	5428147-1	Patent No. 5428147

ALIGNMENTS

RESULT 1
US-09-221-017B-726
; Sequence 726, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998

Not present

```
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ. ID NO: 726:
SEQUENCE CHARACTERISTICS:
LENGTH: 1974 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...1974
US-09-221-017B-726

Alignment Scores:
Pred. No.: 0.605      Length: 1974
Score: 54.00         Matches: 10
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 100.00%            Indels: 0
DB: 4                  Gaps: 0

US-10-008-355-25 (1-10) x US-09-221-017B-726 (1-1974)

QY 1 ThGlyGlyAsnSerGlySerProValPhe 10
|||||
Db 1524 ACGGCGGTAACTCGGGTACCCCGTATTC 1553

RESULT 2
US-09-134-001C-892
Sequence 892, Application US/09134001C
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT FILING DATE: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 892
LENGTH: 936
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-892

Alignment Scores:
Pred. No.: 2.17      Length: 936
Score: 49.00         Matches: 9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 90.74%            Indels: 0
DB: 4                  Gaps: 0

US-10-008-355-25 (1-10) x US-09-134-001C-892 (1-936)

QY 2 GlyGlyAsnSerGlySerProValPhe 10
|||||
Db 781 GGTGAACACTGTGATCTCCAGTATTT 807
```

```
RESULT 3
US-09-221-017B-1045
Sequence 1045, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ. ID NO: 1045:
SEQUENCE CHARACTERISTICS:
LENGTH: 2384 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...2384
US-09-221-017B-1045

Alignment Scores:
Pred. No.: 9.43      Length: 2384
Score: 48.00         Matches: 9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 88.89%            Indels: 0
DB: 4                  Gaps: 0

US-10-008-355-25 (1-10) x US-09-221-017B-1045 (1-2384)

QY 1 ThGlyGlyAsnSerGlySerProVal 9
|||||
```

Db 420 ACCGGCGCACTCAGCAGTCGGGTC 446

RESULT 4

US-09-071-035-427

Sequence 427, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gail H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 427:

SEQUENCE CHARACTERISTICS:

LENGTH: 758 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-427

Alignment Scores:

Pred. No.: 14.2 Length: 758

Score: 44.00 Matches: 7

Percent Similarity: 90.00% Conservative: 2

Best Local Similarity: 70.00% Mismatches: 1

Query Match: 81.48% Indels: 0

DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-071-035-427 (1-758)

OY 1 ThrlGlylAsnserGlyserProvalPhe 10

||||||| |||||||

Db 575 ACCGGCGTCATCTGTCACCAATCTAT 604

RESULT 5

US-09-071-035-425

Sequence 425, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gail H. Choi

TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 425:

SEQUENCE CHARACTERISTICS:

LENGTH: 888 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-09-071-035-425

Alignment Scores:

Pred. No.: 17 Length: 888

Score: 44.00 Matches: 7

Percent Similarity: 90.00% Conservative: 2

Best Local Similarity: 70.00% Mismatches: 1

Query Match: 81.48% Indels: 0

DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-071-035-425 (1-888)

OY 1 ThrlGlylAsnserGlyserProvalPhe 10

||||||| |||||||

Db 703 ACCGGCGTCATCTGTCACCAATCTAT 732

RESULT 6

US-08-035-634-1

Sequence 1, Application US/08035634

Patent No. 5459064

GENERAL INFORMATION:

APPLICANT: SHIONOGI & CO., LTD.

TITLE OF INVENTION: A No. 5459064el Protease

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster

STREET: 545 Middlefield Road, Suite 200

CITY: Menlo Park

STATE: California

COUNTRY: USA

ZIP: 94025

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: EPSON PC-286 Book type laptop

OPERATING SYSTEM: MS-DOS 2.11

SOFTWARE: Wordstar 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/035,634

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/782,372

FILING DATE: 24-OCT-1991

APPLICATION NUMBER: Japanese Patent

APPLICATION NUMBER: Application No. 5459064 2-288110

FILING DATE: October 24, 1990

ATTORNEY/AGENT INFORMATION:

NAME: BOZICEVIC, KARL

REGISTRATION NUMBER: 28,807
REFERENCE/DOCKET NUMBER: 29900-20298.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 677-7000
TELEFAX: (415) 677-7522
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1448 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Bacillus licheniformis
STRAIN: ATCC NO. 14580
FEATURE:
NAME/KEY: coding sequence
LOCATION: 323 to 1270
IDENTIFICATION METHOD: by experiment
NAME/KEY: signal peptide
LOCATION: 323 to 604
IDENTIFICATION METHOD: by experiment
NAME/KEY: mature peptide
LOCATION: 605 to 1270
IDENTIFICATION METHOD: by experiment
OTHER INFORMATION: xaa at -94 position of amino acid
OTHER INFORMATION: sequence: formyl methionine
US-08-035-634-1

Alignment Scores:
Pred. No.: 44.7 Length: 1448
Score: 43.00 Matches: 8
Percent Similarity: 88.89% Conservative: 0
Best Local Similarity: 88.89% Mismatches: 1
Query Match: 79.63% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x US-08-035-634-1 (1-1448)

QY 2 GlyGlyAsnSerGlySerProValPhe 10

DB 1094 GGAGGACAAAGCGCTTCACCGGTATTC 1120

RESULT 7

US-08-998-416-856/C

Sequence 856, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jurgen

APPLICANT: Knechtley, Philipp

APPLICANT: Reibschung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 6239264artis Corporation

STREET: 3054 Cornwalls Road

CITY: Research Triangle Park

STATE: No. 6239264th Carolina

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416

FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 856:
SEQUENCE CHARACTERISTICS:
LENGTH: 743 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1539UP
US-08-998-416-856

Alignment Scores:
Pred. No.: 75.3 Length: 743
Score: 40.00 Matches: 6
Percent Similarity: 80.00% Conservative: 2
Best Local Similarity: 60.00% Mismatches: 2
Query Match: 74.07% Indels: 0
DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-08-998-416-856 (1-743)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10

DB 576 ACCGGTGGATCTGGAGACCTATATAT 547

RESULT 8

US-08-626-169-2

Sequence 2, Application US/08626169

Patent No. 5861248

GENERAL INFORMATION:

APPLICANT: Russell, David W.

APPLICANT: Thigpen, Anice E.

TITLE OF INVENTION: BIOMARKERS FOR DETECTION, DIAGNOSIS

TITLE OF INVENTION: AND PROGNOSIS OF PROSTATE CANCER

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: United States

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/626,169

FILING DATE: Concurrently Herewith

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Corider, Timothy S.

REGISTRATION NUMBER: 38,414

REFERENCE/DOCKET NUMBER: UROC:007

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000

TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 240 base pairs


```

: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
US-08-626-169-2
Alignment Scores:
Pred. No.: 32.5 Length: 240
Score: 39.00 Matches: 7
Percent Similarity: 77.78% Conservative: 0
Best Local Similarity: 77.78% Mismatches: 2
Query Match: 72.22% Indels: 0
DB: 2 Gaps: 0
US-10-008-355-25 (1-10) x US-08-626-169-2 (1-240)
QY 2 GGGGGAACGTGGGGGAACCTGTGTTT 10
DB 6 GGGGGAACGTGGGGGAACCTGTGTTT 32
RESULT 9
US-09-164-907-2
: Sequence 2, Application US/09164907A
: Patent No. 6090559
: GENERAL INFORMATION:
: APPLICANT: RUSSELL, DAVID W.
: APPLICANT: THIGPEN, ANICE E.
: TITLE OF INVENTION: BIOMARKERS FOR DETECTION, DIAGNOSIS AND PROGNOSIS OF
: TITLE OF INVENTION: PROSTATE CANCER
: FILE REFERENCE: UROC:021
: CURRENT APPLICATION NUMBER: US/09/164,907A
: CURRENT FILING DATE: 1998-10-01
: EARLIER APPLICATION NUMBER: 08/626,169
: EARLIER FILING DATE: 1996-03-29
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 2
: LENGTH: 240
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-164-907-2
Alignment Scores:
Pred. No.: 32.5 Length: 240
Score: 39.00 Matches: 7
Percent Similarity: 77.78% Conservative: 0
Best Local Similarity: 77.78% Mismatches: 2
Query Match: 72.22% Indels: 0
DB: 3 Gaps: 0
US-10-008-355-25 (1-10) x US-09-164-907-2 (1-240)
QY 2 GGGGGAACGTGGGGGAACCTGTGTTT 10
DB 6 GGGGGAACGTGGGGGAACCTGTGTTT 32
RESULT 10
US-07-723-002C-5
: Sequence 5, Application US/07723002C
: Patent No. 5447862
: GENERAL INFORMATION:
: APPLICANT: Heim, Jutta
: APPLICANT: Meyhack, Bernd
: APPLICANT: Gysler, Christof
: APPLICANT: Visser, Jacob
: APPLICANT: Kester, Hermanus Cornelis Maria
: TITLE OF INVENTION: No. 5447862el Expression System
: NUMBER OF SEQUENCES: 19
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: New York
: COUNTRY: USA
```

```

: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/723,002C
: FILING DATE: 28-JUL-1991
: CLASSIFICATION: 435
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: GB 8702475
: FILING DATE: 04-FEB-1987
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 150,880
: FILING DATE: 29-JAN-1988
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: GB 8818046.8
: FILING DATE: 28-JUL-1988
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: GB 8914666.6
: FILING DATE: 26-JUN-1989
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 384,898
: FILING DATE: 24-JUL-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Foley, Shawn P.
: REGISTRATION NUMBER: 33,071
: REFERENCE/DOCKET NUMBER: 4-16317/+CIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919)541-8615
: TELEFAX: (919)541-8689
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2774 base pairs.
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Aspergillus niger N400
: IMMEDIATE SOURCE:
: CLONE: Plasmid pGW830 (DSM 4389)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(1134..1337, 1400..1543, 1600..1725, 1783
: LOCATION: ..2112, 2170..2502)
: OTHER INFORMATION: /transl_except= (pos: 1541 .. 1543, aa: Tyr)
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 1134..1193
: FEATURE:
: NAME/KEY: exon
: LOCATION: 1194..1337
: FEATURE:
: NAME/KEY: intron
: LOCATION: 1338..1399
: FEATURE:
: NAME/KEY: exon
: LOCATION: 1400..1542
: FEATURE:
: NAME/KEY: intron
: LOCATION: 1543..1598
: FEATURE:
: NAME/KEY: exon
: LOCATION: 1599..1725
: FEATURE:
: NAME/KEY: intron
: LOCATION: 1726..1782
: FEATURE:
: NAME/KEY: exon
```

```
LOCATION: 1783..2112
FEATURE:
NAME/KEY: Intron
LOCATION: 2113..2169
FEATURE:
NAME/KEY: exon
LOCATION: 2170..2502
FEATURE:
NAME/KEY: promoter
LOCATION: 1..1133
FEATURE:
NAME/KEY: terminator
LOCATION: 2506..2774
US-07-723-002C-5

Alignment Scores:
Pred. No.: 500
Score: 39.00
Percent Similarity: 80.00%
Best Local Similarity: 70.00%
Query Match: 72.22%
DB: 1
Gaps: 0

US-10-008-355-25 (1-10) x US-07-723-002C-5 (1-2774)
Oy 1 ThrGlyAsnSerGlySerProValPhe 10
||||||| ||| |||||||:::
Db 1236 ACTGGCGGTGGACGCTTCCCGCTCTAT 1265

RESULT 11
US-09-222-575-123
Sequence 123, Application US/09222575
Patent No. 6387697
GENERAL INFORMATION:
APPLICANT: Inqiu, Jiang
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: Compositions for the Treatment and Diagnosis of Breast Cancer
TITLE OF INVENTION: and Methods for Their Use
FILE REFERENCE: 210121.470
CURRENT APPLICATION NUMBER: US/09/222,575
CURRENT FILING DATE: 1998-12-28
NUMBER OF SEQ ID NOS: 174
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 123
LENGTH: 531
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: modified_base
LOCATION: (152)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (373)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (482)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (494)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (496)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (502)
OTHER INFORMATION: Where n is a, c, g or t
US-09-222-575-123

Alignment Scores:
Pred. No.: 120
Score: 38.00
Length: 531
Matches: 6
```

```
Percent Similarity: 77.78%
Best Local Similarity: 66.67%
Query Match: 70.37%
DB: 4
Gaps: 0

US-10-008-355-25 (1-10) x US-09-222-575-123 (1-531)
Oy 2 GlyGlyAsnSerGlySerProValPhe 10
||||| |||||||:::
Db 241 GGGGAGGTGGGGAACCTATTTTT 267

RESULT 12
US-07-624-313-3/c
Sequence 3, Application US/07624313
Patent No. 5250411
GENERAL INFORMATION:
APPLICANT: Ayyanathan, K.
APPLICANT: Bhat, P.
APPLICANT: Datta, S.
APPLICANT: Francis, V.S.N.K.
APPLICANT: Padmanaban, G.
APPLICANT: Srivivasa, H.
TITLE OF INVENTION: NEW ANALYSIS METHOD
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESS: White & Case
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/624,313
FILING DATE: 19901204
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 8904100-8
FILING DATE: 05-DEC-1989
ATTORNEY/AGENT INFORMATION:
NAME: Ryan, John W.
REGISTRATION NUMBER: 33,771
REFERENCE/DOCKET NUMBER: 1103326-811
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-819-8515
TELEFAX: 212-354-8113
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 663 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium vivax
IMMEDIATE SOURCE:
CLONE: PARC 1153
US-07-624-313-3

Alignment Scores:
Pred. No.: 154
Score: 38.00
Percent Similarity: 77.78%
Best Local Similarity: 77.78%
Query Match: 70.37%
DB: 1
Gaps: 0
Length: 663
Matches: 7
Conservative: 0
Mismatch: 2
Indels: 0
```

US-10-008-355-25 (1-10) x US-07-624-313-3 (1-663)

OY 2 GLYGLYAsnSerGlySerProValPhe 10

DB 220 GGAGGAACAGTGTGGAAACCTGGATT 194

RESULT 13

US-09-061-709-8

Sequence 8, Application US/09061709B

Patent No. 6297364

GENERAL INFORMATION:

APPLICANT: Chen, Yao-Tseng

APPLICANT: Gure, Ali

APPLICANT: Tsang, Solam

APPLICANT: Stockert, Elisabeth

APPLICANT: Jager, Elke

APPLICANT: Knuth, Alexander

APPLICANT: Old, Lloyd J.

TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated

TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof

FILE REFERENCE: LUD 5538

CURRENT APPLICATION NUMBER: US/09/061,709B

CURRENT FILING DATE: 1998-04-17

NUMBER OF SEQ ID NOS: 8

SEQ ID NO 8

LENGTH: 3283

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

US-09-061-709-8

Alignment Scores:

Pred. No.: 922 Length: 3283

Score: 38.00 Matches: 7

Percent Similarity: 90.00% Conservative: 2

Best Local Similarity: 70.00% Mismatches: 1

Query Match: 70.37% Indels: 0

DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-061-709-8 (1-3283)

OY 1 ThrGLYGLYAsnSerGlySerProValPhe 10

DB 2406 ACTCAGGCGAGTGTGGATCAGCTGTGTAT 2435

RESULT 14

US-09-061-709-6

Sequence 6, Application US/09061709B

Patent No. 6297364

GENERAL INFORMATION:

APPLICANT: Chen, Yao-Tseng

APPLICANT: Gure, Ali

APPLICANT: Tsang, Solam

APPLICANT: Stockert, Elisabeth

APPLICANT: Jager, Elke

APPLICANT: Knuth, Alexander

APPLICANT: Old, Lloyd J.

TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated

TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof

Percent Similarity: 90.00% Conservative: 2

Best Local Similarity: 70.00% Mismatches: 1

Query Match: 70.37% Indels: 0

DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-061-709-6 (1-3412)

OY 1 ThrGLYGLYAsnSerGlySerProValPhe 10

DB 2535 ACTCAGGCGAGTGTGGATCAGCTGTGTAT 2564

RESULT 15

US-08-225-224-31/c

Sequence 31, Application US/08225224

Patent No. 5635599

GENERAL INFORMATION:

APPLICANT: PASTAN, Ira

APPLICANT: KREITMAN, Robert J.

TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND

TITLE OF INVENTION: CIRCULARLY PERMUTED FUSION PROTEINS

NUMBER OF SEQUENCES: 57

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Townsend and Townsend Kourie and Crew

STREET: Stewart Street Tower, One Market Plaza

CITY: San Francisco

STATE: California

COUNTRY: US

ZIP: 94105-1493

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/225,224

FILING DATE: 8-Apr-1994

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Ellen L.

REGISTRATION NUMBER: 32,762

REFERENCE/DOCKET NUMBER: 15280-193

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043

SEQUENCE CHARACTERISTICS:

INFORMATION FOR SEQ ID NO: 31:

LENGTH: 39 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (primer)

US-08-225-224-31

Alignment Scores:

Pred. No.: 9.93 Length: 39

Score: 37.00 Matches: 6

Percent Similarity: 87.50% Conservative: 1

Best Local Similarity: 75.00% Mismatches: 1

Query Match: 68.52% Indels: 0

DB: 1 Gaps: 0

US-10-008-355-25 (1-10) x US-08-225-224-31 (1-39)

OY 1 ThrGLYGLYAsnSerGlySerPro 8

DB 24 ACCGAGGTAAACGTGGGACCT 1

Search completed: December 20, 2002, 14:11:50

JOB time : 42.4211 secs

Alignment Scores:

Pred. No.: 962 Length: 3412

Score: 38.00 Matches: 7

Percent Similarity: 90.00% Conservative: 2

Best Local Similarity: 70.00% Mismatches: 1

Query Match: 70.37% Indels: 0

DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-061-709-8 (1-3283)

OY 1 ThrGLYGLYAsnSerGlySerProValPhe 10

DB 2406 ACTCAGGCGAGTGTGGATCAGCTGTGTAT 2435

RESULT 14

US-09-061-709-6

Sequence 6, Application US/09061709B

Patent No. 6297364

GENERAL INFORMATION:

APPLICANT: Chen, Yao-Tseng

APPLICANT: Gure, Ali

APPLICANT: Tsang, Solam

APPLICANT: Stockert, Elisabeth

APPLICANT: Jager, Elke

APPLICANT: Knuth, Alexander

APPLICANT: Old, Lloyd J.

TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated

TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof

FILE REFERENCE: LUD 5538

CURRENT APPLICATION NUMBER: US/09/061,709B

CURRENT FILING DATE: 1998-04-17

NUMBER OF SEQ ID NOS: 8

SEQ ID NO 8

LENGTH: 3283

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

US-09-061-709-8

Alignment Scores:

Pred. No.: 922 Length: 3283

Score: 38.00 Matches: 7

Percent Similarity: 90.00% Conservative: 2

Best Local Similarity: 70.00% Mismatches: 1

Query Match: 70.37% Indels: 0

DB: 4 Gaps: 0

US-10-008-355-25 (1-10) x US-09-061-709-8 (1-3283)

OY 1 ThrGLYGLYAsnSerGlySerProValPhe 10

DB 2406 ACTCAGGCGAGTGTGGATCAGCTGTGTAT 2435

RESULT 14

US-09-061-709-6

Sequence 6, Application US/09061709B

Patent No. 6297364

GENERAL INFORMATION:

APPLICANT: Chen, Yao-Tseng

APPLICANT: Gure, Ali

APPLICANT: Tsang, Solam

APPLICANT: Stockert, Elisabeth

APPLICANT: Jager, Elke

APPLICANT: Knuth, Alexander

APPLICANT: Old, Lloyd J.

TITLE OF INVENTION: Isolated Nucleic Acid Molecules Encoding Cancer Associated

TITLE OF INVENTION: Antigen, The Antigens Per Se, And Uses Thereof

FILE REFERENCE: LUD 5538

CURRENT APPLICATION NUMBER: US/09/061,709B

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:17:34 ; Search time 40 Seconds

(without alignments)
99.214 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TGCNGSPVF 10

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 356696 segs, 198428768 residues

Total number of hits satisfying chosen parameters: 713392

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=11h
-Q/cgcn2.1/USPTO-SPOOL/US10008355/runat.17122002.112339.14635/app-query.fasta.1.398
-DB=Published_Applications_NA -QPM=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10008355.@CGN.1.1.21/runat.17122002.112339.14635
-NCPu=6 -ICPU=3 -NO_XLPEXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1: /cgcn2.6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgcn2.6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgcn2.6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgcn2.6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgcn2.6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgcn2.6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgcn2.6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgcn2.6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgcn2.6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgcn2.6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgcn2.6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgcn2.6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgcn2.6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgcn2.6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	54	100.0	2139	9	US-10-008-355-1
2	44	81.5	4951	10	US-09-070-927A-261
3	43	79.6	534	10	US-09-974-300-337
4	40	74.1	450	10	US-09-867-701-1705

C 5	40	74.1	7441	10	US-09-764-878-243	Sequence 243, App
C 6	39	72.2	235	10	US-09-960-352-5672	Sequence 5672, Ap
C 7	39	72.2	584	10	US-09-864-761-9954	Sequence 9954, Ap
8	39	72.2	1775	9	US-10-001-887-75	Sequence 75, Appl
9	39	72.2	2109	9	US-09-764-868-111	Sequence 111, App
10	39	72.2	3255	9	US-10-025-380-663	Sequence 683, App
11	39	72.2	3255	10	US-09-922-217-683	Sequence 683, App
12	39	72.2	3255	10	US-09-833-263-683	Sequence 683, App
13	38	70.4	218	10	US-09-815-343-1415	Sequence 1415, Ap
14	38	70.4	222	10	US-09-815-343-676	Sequence 676, App
15	38	70.4	222	10	US-09-815-343-1285	Sequence 1285, Ap
16	38	70.4	243	10	US-09-815-343-614	Sequence 614, App
17	38	70.4	253	10	US-09-815-343-729	Sequence 729, App
18	38	70.4	454	10	US-09-770-444-591	Sequence 591, App
C 19	38	70.4	474	10	US-09-867-701-5230	Sequence 5230, Ap
C 20	38	70.4	491	10	US-09-783-590-6501	Sequence 6501, Ap
21	38	70.4	508	10	US-09-783-590-4015	Sequence 4015, Ap
22	38	70.4	531	10	US-09-604-287A-123	Sequence 123, App
23	38	70.4	531	10	US-09-339-338-123	Sequence 123, App
24	38	70.4	531	12	US-10-007-805-123	Sequence 123, App
25	38	70.4	1114	10	US-09-808-701-14	Sequence 14, Appl
26	38	70.4	2290	10	US-09-764-864-332	Sequence 332, App
27	38	70.4	3283	10	US-09-899-651-8	Sequence 8, Appl1
28	38	70.4	3412	10	US-09-899-651-6	Sequence 6, Appl1
29	38	70.4	3694	10	US-09-764-864-329	Sequence 329, App
C 30	37.5	69.4	400	10	US-09-960-352-14943	Sequence 14943, A
C 31	37	68.5	222	10	US-09-864-761-15889	Sequence 32993, A
C 32	37	68.5	557	10	US-09-864-761-15889	Sequence 15889, A
33	37	68.5	720	10	US-09-815-242-4296	Sequence 4296, Ap
34	37	68.5	765	10	US-09-815-242-8179	Sequence 8179, Ap
35	37	68.5	2000	9	US-09-938-842A-3053	Sequence 3053, Ap
C 36	37	68.5	2205	10	US-09-822-830A-458	Sequence 458, App
C 37	37	68.5	51552	10	US-09-733-294A-30	Sequence 30, Appl
C 38	36	66.7	191	10	US-09-864-761-24491	Sequence 24491, A
39	36	66.7	297	10	US-09-783-590-10996	Sequence 10996, A
C 40	36	66.7	372	10	US-09-560-863-753	Sequence 753, App
41	36	66.7	393	10	US-09-960-352-3735	Sequence 3735, App
42	36	66.7	441	10	US-09-960-352-9622	Sequence 9622, App
C 43	36	66.7	529	10	US-09-864-761-7787	Sequence 7787, Ap
44	36	66.7	1041	10	US-09-751-299-1	Sequence 1, Appl1
45	36	66.7	1142	9	US-09-764-868-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-10-008-355-1
; Sequence 1, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235, 00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-1

Alignment Scores:
Pred. No.: 0.363
Score: 54.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Length: 2139
Matches: 10
Conservative: 0
Mismatch: 0
Indels: 0

DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x US-10-008-355-1 (1-2139)

Qy 1 ThGlyGlyAsnSerGlySerProValPhe 10
|||||

Db 1930 .ACGGCGGTACATCGCGTACCCCGATATTC 1959

RESULT 2

US-09-070-927A-261/C

; Sequence 261, Application US/09070927A

; Patent No. US20020120116A1

; GENERAL INFORMATION:

; APPLICANT: Charles A. Kunsch
Patrick J. Dillon

; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

; NUMBER OF SEQUENCES: 982

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: Maryland

; COUNTRY: USA

; ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 261:

SEQUENCE CHARACTERISTICS:

LENGTH: 4951 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 261:

US-09-070-927A-261

Alignment Scores:

Pred. No.: 71.5 Length: 4951

Score: 44.00 Matches: 7

Percent Similarity: 90.00% Conservative: 2

Best Local Similarity: 70.00% Mismatches: 1

Query Match: 81.48% Indels: 0

DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-070-927A-261 (1-4951)

Qy 1 ThGlyGlyAsnSerGlySerProValPhe 10
|||||

Db 695 ACCGGCGGTACATCTGTTCCACCAATCTAT 666

RESULT 3

US-09-974-300-327

; Sequence 327, Application US/09974300

; Patent No. US20020146721A1

; GENERAL INFORMATION:

; APPLICANT: Berka, Randy M.

; APPLICANT: Clausen, Ib Groth

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene

; FILE REFERENCE: 10085,500-US

; CURRENT APPLICATION NUMBER: US/09/974,300

; CURRENT FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: 09/680,598

; PRIOR FILING DATE: 2000-10-06

; PRIOR APPLICATION NUMBER: 60/279,526

; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 327

LENGTH: 534

TYPE: DNA

ORGANISM: Bacillus licheniformis

US-09-974-300-327

Alignment Scores:

Pred. No.: 8.61 Length: 534

Score: 43.00 Matches: 8

Percent Similarity: 88.89% Conservative: 0

Best Local Similarity: 88.89% Mismatches: 1

Query Match: 79.63% Indels: 0

DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-974-300-327 (1-534)

Qy 2 GlyGlyAsnSerGlySerProValPhe 10
|||||

Db 357 GGAGGACAAAGCGGTTCACCGGATATTC 383

RESULT 4

US-09-867-701-1705

; Sequence 1705, Application US/09867701

; Patent No. US20020132237A1

; GENERAL INFORMATION:

; APPLICANT: Aglate, Paul A.

; APPLICANT: Jones, Robert

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701

; CURRENT FILING DATE: 2001-05-29

; NUMBER OF SEQ ID NOS: 10912

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1705

LENGTH: 450

TYPE: DNA

ORGANISM: Homo sapien

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(450)

OTHER INFORMATION: n = A,T,C or G

US-09-867-701-1705

Alignment Scores:

Pred. No.: 25.9 Length: 450

Score: 40.00 Matches: 7

Percent Similarity: 87.50% Conservative: 0

Best Local Similarity: 87.50% Mismatches: 1

Query Match: 74.07% Indels: 0

DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x US-09-867-701-1705 (1-450)

Qy 1 ThGlyGlyAsnSerGlySerPro 8
|||||

Db 82 ACAGAGTAATTCTGCAGCCCC 105

RESULT 5

US-09-764-878-243/C

Sequence 243, Application US/09764878

Patent No. US20020090615A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PA121

CURRENT APPLICATION NUMBER: US/09/764,878

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 428

SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 243

LENGTH: 7441

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-878-243

Alignment Scores:

Pred. No.: 643

Score: 40.00

Percent Similarity: 100.00%

Best Local Similarity: 87.50%

Query Match: 74.07%

DB: 10

Gaps: 0

US-10-008-355-25 (1-10) x US-09-764-878-243 (1-7441)

OY 2 GLYGLYAsnserGlyserProval 9

|||||:|||||

Db 2023 GCGGTACACAGCAGCCCTGTG 2000

RESULT 6

US-09-960-352-5672/C

Sequence 5672, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warren, Wesley C.

APPLICANT: Tao, Nengbing

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

FILE REFERENCE: 16511.006/37-21(10238)C

CURRENT APPLICATION NUMBER: US/09/960,352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 5672

LENGTH: 235

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 24-LIB34-067-Q1-E1-F7

US-09-960-352-5672

Alignment Scores:

Pred. No.: 19

Score: 39.00

Percent Similarity: 77.78%

Best Local Similarity: 72.22%

Query Match: 72.22%

DB: 10

Gaps: 0

US-10-008-355-25 (1-10) x US-09-960-352-5672 (1-235)

OY 2 GLYGLYAsnserGlyserProvalphe 10

||||| ||||| |||||

Db 211 GGAGGCAAGTCAGGTGTCAGTAATTC 185

RESULT 7

US-09-864-761-9954/C

Sequence 9954, Application US/09864761

Patent No. US20020048763A1

GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.

APPLICANT: Hanzel, David K.

APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

FILE REFERENCE: Aemica-X-1

CURRENT APPLICATION NUMBER: US/09/864,761

CURRENT FILING DATE: 2001-05-23

Prior APPLICATION NUMBER: US 60/180,312

Prior FILING DATE: 2000-02-04

Prior APPLICATION NUMBER: US 60/207,456

Prior FILING DATE: 2000-05-26

Prior APPLICATION NUMBER: US 09/632,366

Prior FILING DATE: 2000-08-03

Prior APPLICATION NUMBER: GB 24263.6

Prior FILING DATE: 2000-10-04

Prior APPLICATION NUMBER: US 60/236,359

Prior FILING DATE: 2000-09-27

Prior APPLICATION NUMBER: PCT/US01/00666

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00667

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00664

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00669

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00665

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00668

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00663

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00662

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00661

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: PCT/US01/00670

Prior FILING DATE: 2001-01-30

Prior APPLICATION NUMBER: US 60/234,687

Prior FILING DATE: 2000-09-21

Prior APPLICATION NUMBER: US 09/608,408

Prior FILING DATE: 2000-06-30

Prior APPLICATION NUMBER: US 09/774,203

Prior FILING DATE: 2001-01-29

NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 9954

LENGTH: 584

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: MAP TO AC007463.3

OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4

OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5

US-09-864-761-9954

Alignment Scores:

Pred. No.: 53.8

Score: 39.00

Percent Similarity: 100.00%

Best Local Similarity: 75.00%

Query Match: 72.22%

DB: 10

Gaps: 0

US-10-008-355-25 (1-10) x US-09-864-761-9954 (1-584)

OY 3 GLYAsnserGlyserProvalphe 10

|||||:|||||:|||||

Db 160 GGGAAGCTGTGAGCAGCCCAATATT 137

```

QY      2 GlyGlyAsnSerGlySerProValPhe 10
      |||||

```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525


```

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 683
LENGTH: 3255
TYPE: DNA
ORGANISM: Homo sapiens
US-09-922-217-683

Alignment Scores:
Pred. No.:          384      Length:    3255
Score:              39.00     Matches:    7
Percent Similarity: 77.78%   Conservative: 0
Best local Similarity: 77.78% Mismatches:    2
Query Match:        72.22%   Indels:     0
DB:                 10       Gaps:      0

US-10-008-355-25 (1-10) x US-09-922-217-683 (1-3255)
OY      2 glyglyasnserglyserprovalphe 10
||||||| ||| |||||||||
Db      2988 GGGGGGAACGTGGGGAACCTGTGTTT 3014

RESULT 12
Sequence 683, Application US/09833263
Patent No. US20020110547A1
GENERAL INFORMATION:
APPLICANT: Wang, Aijun
APPLICANT: Clapper, Jonathan D.
APPLICANT: Stolk, John A.
APPLICANT: Meagher, Madeleine J.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.471c12
CURRENT APPLICATION NUMBER: US/09/833,263
CURRENT FILING DATE: 2001-04-10
NUMBER OF SEQ. ID NOS.: 1093
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 683
LENGTH: 3255
TYPE: DNA
ORGANISM: Homo saplen
US-09-833-263-683

Alignment Scores:
Pred. No.:          384      Length:    3255
Score:              39.00     Matches:    7
Percent Similarity: 77.78%   Conservative: 0
Best local Similarity: 77.78% Mismatches:    2
Query Match:        72.22%   Indels:     0
DB:                 10       Gaps:      0

US-10-008-355-25 (1-10) x US-09-833-263-683 (1-3255)
OY      2 glyglyasnserglyserprovalphe 10
||||||| ||| |||||||||
Db      2988 GGGGGGAACGTGGGGAACCTGTGTTT 3014

RESULT 13
US-09-815-343-1415
Sequence 1415, Application US/09815343
Patent No. US2001005596A1
GENERAL INFORMATION:
APPLICANT: Meagher, Madeleine
APPLICANT: Xu, Jiangchun
APPLICANT: King, Gordon E.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.504
CURRENT APPLICATION NUMBER: US/09/815,343
CURRENT FILING DATE: 2001-03-22
NUMBER OF SEQ. ID NOS.: 1556
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1415

```

```

?      LENGTH: 218
?      TYPE: DNA
?      ORGANISM: Homo sapien
US-09-815-343-1415

Alignment Scores:
Pred. No.:          26 8          Length:          218
Score:             38.00         Matches:          7
Percent Similarity: 90.00%       Conservative:    2
Best Local Similarity: 70.00%    Mismatches:     1
Query Match:       70.37%        Indels:         0
DB:               10            Gaps:             0

US-10-008-355-25 (1-10) x US-09-815-343-1415 (1-218)
Qy      1 ThrglyGlyAsnserGlySerProvalPhe 10
      ||| |||:|||||:|||||:|||||:|
Db      96 ACTCAGCGCAGTTCGTGATCCTGCTGTAT 125

RESULT 14
US-09-815-343-676
? Sequence 676, Application US/09815343
? Patent No. US20010055596A1
? GENERAL INFORMATION:
? APPLICANT: Meagher, Madeleine
? APPLICANT: Xu, Jiangchun
? APPLICANT: King, Gordon E.
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
? TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
? FILE REFERENCE: 210121.504
? CURRENT APPLICATION NUMBER: US/09/815,343
? CURRENT FILING DATE: 2001-03-22
? NUMBER OF SEQ ID NOS: 1556
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 676
? LENGTH: 222
? TYPE: DNA
? ORGANISM: Homo sapien
US-09-815-343-676

Alignment Scores:
Pred. No.:          27 4          Length:          222
Score:             38.00         Matches:          7
Percent Similarity: 90.00%       Conservative:    2
Best Local Similarity: 70.00%    Mismatches:     1
Query Match:       70.37%        Indels:         0
DB:               10            Gaps:             0

US-10-008-355-25 (1-10) x US-09-815-343-676 (1-222)
Qy      1 ThrglyGlyAsnserGlySerProvalPhe 10
      ||| |||:|||||:|||||:|||||:|
Db      96 ACTCAGCGCAGTTCGTGATCCTGCTGTAT 125

RESULT 15
US-09-815-343-1285
? Sequence 1285, Application US/09815343
? Patent No. US20010055596A1
? GENERAL INFORMATION:
? APPLICANT: Meagher, Madeleine
? APPLICANT: Xu, Jiangchun
? APPLICANT: King, Gordon E.
? TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
? TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
? FILE REFERENCE: 210121.504
? CURRENT APPLICATION NUMBER: US/09/815,343
? CURRENT FILING DATE: 2001-03-22
? NUMBER OF SEQ ID NOS: 1556
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 1285
? LENGTH: 222
? TYPE: DNA
? ORGANISM: Homo sapien

```

US-09-815-343-1285

Alignment Scores:

Pred. No.:	27.4	Length:	222
Score:	38.00	Matches:	7
Percent Similarity:	90.00%	Conservative:	2
Best Local Similarity:	70.00%	Mismatches:	1
Query Match:	70.37%	Indels:	0
DB:	10	Gaps:	0

US-10-008-355-25 (1-10) x US-09-815-343-1285 (1-222)

Qy	1	ThrGlyGlyAsnSerGlySerProValPhe	10
Db	96	ACTCACGCGCAGTTCTGTGATCACCTGTGTAT	125

Search completed: December 20, 2002, 17:10:39
Job time : 45 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 09:34:23 : Search time 17.3684 Seconds
(without alignments)
16.940 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TCGNGSPVF 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	213	3	US-08-523-373-22 Sequence 22, App1
2	54	100.0	214	3	US-08-523-373-23 Sequence 23, App1
3	54	100.0	215	3	US-08-523-373-24 Sequence 24, App1
4	54	100.0	344	1	US-08-523-373-5 Sequence 5, App1
5	54	100.0	344	3	US-08-523-373-5 Sequence 5, App1
6	54	100.0	392	3	US-08-523-373-6 Sequence 6, App1
7	54	100.0	532	1	US-08-523-373-7 Sequence 7, App1
8	54	100.0	532	3	US-08-523-373-7 Sequence 7, App1
9	54	100.0	537	1	US-08-523-373-7 Sequence 7, App1
10	49	90.7	311	4	US-09-134-001C-3729 Sequence 3729, App1
11	48	88.9	15	2	US-08-332-562A-67 Sequence 67, App1
12	44	81.5	252	4	US-09-071-035-428 Sequence 428, App1
13	44	81.5	284	4	US-09-071-035-428 Sequence 428, App1
14	43	79.6	222	1	US-08-090-048-1 Sequence 1, App1
15	43	79.6	222	2	US-08-292-550-1 Sequence 1, App1
16	43	79.6	222	2	US-07-927-661A-1 Sequence 1, App1
17	42	77.8	418	4	US-09-342-653-7 Sequence 7, App1
18	39	72.2	379	1	US-07-723-002C-6 Sequence 6, App1
19	36	66.7	379	2	US-08-887-365-36 Sequence 36, App1
20	36	66.7	379	2	US-09-194-905-13 Sequence 13, App1
21	36	66.7	628	2	US-07-952-853-22 Sequence 22, App1
22	36	66.7	628	2	US-08-914-848-22 Sequence 22, App1
23	35	64.8	8	4	US-08-296-791-7 Sequence 7, App1
24	35	64.8	8	4	US-08-296-791-7 Sequence 7, App1
25	35	64.8	8	5	PCT-US95-10661A-7 Sequence 7, App1
26	35	64.8	8	5	PCT-US95-10661A-8 Sequence 8, App1
27	35	64.8	318	4	US-09-060-756-727 Sequence 727, App1

28	35	64.8	334	4	US-09-060-756-728 Sequence 728, App1
29	35	64.8	503	1	US-08-245-294-8 Sequence 8, App1
30	35	64.8	503	1	US-08-474-499-8 Sequence 8, App1
31	35	64.8	503	1	US-08-307-279A-8 Sequence 8, App1
32	35	64.8	503	4	US-09-525-310-8 Sequence 8, App1
33	35	64.8	503	5	PCT-US95-06211-8 Sequence 8, App1
34	35	64.8	550	4	US-08-868-373-8 Sequence 8, App1
35	35	64.8	551	2	US-09-033-537A-1 Sequence 1, App1
36	35	64.8	1394	4	US-08-296-791-2 Sequence 2, App1
37	35	64.8	1394	5	PCT-US95-10661A-2 Sequence 2, App1
38	35	64.8	1507	6	PCT-US95-10661A-2 Patent No. 5268270
39	35	64.8	1541	4	US-08-296-791-3 Sequence 3, App1
40	35	64.8	1541	5	PCT-US95-10661A-3 Sequence 3, App1
41	35	64.8	1545	4	US-08-296-791-4 Sequence 4, App1
42	35	64.8	1545	5	PCT-US95-10661A-4 Sequence 4, App1
43	35	64.8	1702	4	US-08-296-791-5 Sequence 5, App1
44	35	64.8	1702	5	PCT-US95-10661A-5 Sequence 5, App1
45	35	64.8	1848	4	US-08-296-791-6 Sequence 6, App1

ALIGNMENTS

RESULT 1
US-08-523-373-22
; Sequence 22, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohnuye, Kazuhiko
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-523-373-22
Query Match 100.0%; Score 54; DB 3; Length 213;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGPVF 10
|||||
Db 165 TGNSSGPVF 174

RESULT 2

US-08-523-373-23
; Sequence 23, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohsuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373-23

Query Match 100.0%; Score 54; DB 3; Length 214;

Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGPVF 10
|||||
Db 165 TGNSSGPVF 174

RESULT 3

US-08-523-373-24
; Sequence 24, Application US/08523373
; Patent No. 6037145
; GENERAL INFORMATION:
; APPLICANT: Yabuta, Masayuki
; APPLICANT: Ohsuye, Kazuhiro
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

STREET: 699 Prince Street
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314-3187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/523,373
; FILING DATE: 05-SEP-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 6-238595
; FILING DATE: 07-SEP-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 6-296028
; FILING DATE: 07-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-251
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 215 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373-24

Query Match 100.0%; Score 54; DB 3; Length 215;

Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGPVF 10
|||||
Db 165 TGNSSGPVF 174

RESULT 4

US-08-657-192-3
; Sequence 3, Application US/08657192
; Patent No. 5747321
; GENERAL INFORMATION:
; APPLICANT: YABUTA, Masayuki
; APPLICANT: OHSUYE, Kazuhiro
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/657,192
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: JP 7-170086

FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Neuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-657-192-3

Query Match 100.0%; Score 54; DB 1; Length 344;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSNGSPVF 10
|||||
DB 289 TGSNGSPVF 298

RESULT 5
US-08-523-373-5
Sequence 5, Application US/08523373
Patent No. 6037145
GENERAL INFORMATION:
APPLICANT: Yabuta, Masayuki
APPLICANT: Ohsuye, Kazuhito
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/523,373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-238595
FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028
FILING DATE: 07-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Neuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-5

Query Match 100.0%; Score 54; DB 3; Length 344;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSNGSPVF 10
|||||
DB 289 TGSNGSPVF 298

RESULT 6
US-08-523-373-6
Sequence 6, Application US/08523373
Patent No. 6037145
GENERAL INFORMATION:
APPLICANT: Yabuta, Masayuki
APPLICANT: Ohsuye, Kazuhito
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/523,373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-238595
FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028
FILING DATE: 07-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Neuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 392 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-6

Query Match 100.0%; Score 54; DB 3; Length 392;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGSNGSPVF 10
|||||
DB 289 TGSNGSPVF 298

RESULT 7
US-08-657-192-9
Sequence 9, Application US/08657192
Patent No. 5747321
GENERAL INFORMATION:
APPLICANT: YABUTA, Masayuki
APPLICANT: OHSUYE, Kazuhito
TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8

TITLE OF INVENTION: PROTEASES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22113-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/657,192
FILING DATE: 03-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-170086
FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-657-192-9

Query Match 100.0%; Score 54; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10
|||||
DB 289 TCGNSGSPVF 298

RESULT 8
US-08-523-373-7
Sequence 7, Application US/08523373
Patent No. 6037145
GENERAL INFORMATION:
APPLICANT: Yabuta, Masayuki
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22114-3187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/523,373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-238595

FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028
FILING DATE: 07-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-7

Query Match 100.0%; Score 54; DB 3; Length 532;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10
|||||
DB 289 TCGNSGSPVF 298

RESULT 9
US-08-657-192-15
Sequence 15, Application US/08657192
Patent No. 5747321
GENERAL INFORMATION:
APPLICANT: YABUTA, Masayuki
TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22113-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/657,192
FILING DATE: 03-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-170086
FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-657-192-15

Query Match 100.0%; Score 54; DB 1; Length 537;
Best Local Similarity 100.0%; Pred. No. 0.35;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10
|||||
DB 289 TCGNSGSPVF 298

RESULT 10
US-09-134-001C-3729
; Sequence 3729, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3729
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3729

Query Match 90.7%; Score 49; DB 4; Length 311;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGNSSPVF 10
|||||
DB 261 GGNSSPVF 269

RESULT 11
US-08-332-562A-67
; Sequence 67, Application US/08332562A
; Patent No. 5985599
; GENERAL INFORMATION:
; APPLICANT: MCKENZIE, Ian F.C.
; APPLICANT: HOGARTH, Mark P.
; APPLICANT: HIBBS, Margaret L.
; APPLICANT: SCOTT, Bernadette M.
; APPLICANT: BONADONNA, Lisa
; APPLICANT: HULETT, Mark D.
; TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,562A
FILING DATE: 31-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/896,457
FILING DATE: 27-MAY-1992

ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 54270/119/GRRA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-332-562A-67

Query Match 88.9%; Score 48; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.089;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||
DB 5 TCGNSGSPV 13

RESULT 12
US-09-071-035-428
; Sequence 428, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: GILL H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 436
; CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035

CLASSIFICATION:
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8504

INFORMATION FOR SEQ ID NO: 428:
SEQUENCE CHARACTERISTICS:
LENGTH: 252 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-428

Query Match 81.5%; Score 44; DB 4; Length 252;
Best Local Similarity 70.0%; Pred. No. 6.4;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10

DB 192 TGGGSGSPV 201

RESULT 13

US-09-071-035-426
Sequence 426, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 426:
SEQUENCE CHARACTERISTICS:
LENGTH: 284 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-426

Query Match 81.5%; Score 44; DB 4; Length 284;
Best Local Similarity 70.0%; Pred. No. 7.2;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGGSGSPV 10
DB 224 TGGGSGSPV 233

RESULT 14
US-08-090-048-1
Sequence 1, Application US/08090048
Patent No. 5523237
GENERAL INFORMATION:
APPLICANT: Budtz, Peter
TITLE OF INVENTION: PROTEIN PREPARATIONS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 55232370 No. 5523237disk of No. 5523237th America, Inc.
STREET: 405 Lexington Avenue, 62nd Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10174-6201
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/090,048
FILING DATE: 16-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 199/91
FILING DATE: 06-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK92/00036
FILING DATE: 06-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: Agis, Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 3396.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-0298
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-090-048-1

Query Match 79.6%; Score 43; DB 1; Length 222;
Best Local Similarity 88.9%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGGSGSPV 10
DB 164 GGGSGSPV 172

RESULT 15
US-08-292-550-1
Sequence 1, Application US/08292550
Patent No. 5863573
GENERAL INFORMATION:
APPLICANT: Damhann, Claus
APPLICANT: Budtz, Peter
TITLE OF INVENTION: A PROCESS FOR PRODUCING CHEESE
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58635730 No. 5863573disk of No. 5863573th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/292,550
FILING DATE: 18 AUG 1994
CLASSIFICATION: 426
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/927,624
FILING DATE: 31 AUG 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/0068
FILING DATE: 8 MAR 1991
ATTORNEY/AGENT INFORMATION:
NAME: Greg9, Valeta A.

REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 222 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-292-550-1

Query Match 79.6%; Score 43; DB 2; Length 222;
Best Local Similarity 88.9%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGNSGSPVF 10
| | | | | | |
Db 164 GGOSGSPVF 172

Search completed: December 20, 2002, 12:12:18
Job time : 19.3684 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:12:24 : Search time 1521.05 Seconds

(without alignments)
106.476 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TGNSSGSPVF 10

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus.p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US10008355/runat_1122002_112338_14609/app_query.fasta_1.398
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human4.0.cdi -LIST=45
-LOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_ecgn_1.1_761_@runat_1122002_112338_14609 -NCPU=6 -ICPU=3
-NO_XLPTX -NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:*
2: em_estchum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_din:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_fod:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	47	87.0	318	9	AA411980
C 2	47	87.0	426	17	BH400391
C 3	47	87.0	608	10	BB622700
C 4	46	85.2	611	12	BF694219
C 5	45	83.3	371	17	AO699991
C 6	45	83.3	497	14	BO327884
C 7	45	83.3	796	12	BG714149
C 8	45	83.3	914	12	BF128636
C 9	45	83.3	1108	12	BG115436
C 10	44	81.5	237	12	BF893687
C 11	44	81.5	279	10	AM060086
C 12	44	81.5	298	13	BJ099916
C 13	44	81.5	403	9	AA476126
C 14	44	81.5	420	14	W72307
C 15	44	81.5	432	10	AW743569
C 16	44	81.5	447	9	AL117983
C 17	44	81.5	447	9	AL175055
C 18	44	81.5	450	9	AL034929
C 19	44	81.5	491	9	AL034929
C 20	44	81.5	528	13	BI440854
C 21	44	81.5	715	12	BG751121
C 22	44	81.5	768	13	BG918747
C 23	44	81.5	812	17	AO912379
C 24	44	81.5	857	12	BF668889
C 25	44	81.5	873	13	BI695748
C 26	44	81.5	958	14	BO881134
C 27	44	81.5	970	12	BE794754
C 28	44	81.5	3200	11	AK019499
C 29	43	79.6	278	10	BI176886
C 30	43	79.6	377	10	AM308476
C 31	43	79.6	392	10	AM315936
C 32	43	79.6	414	12	BO834820
C 33	43	79.6	422	13	BI830377
C 34	43	79.6	473	12	BG383001
C 35	43	79.6	483	13	BI190590
C 36	43	79.6	584	12	BG834531
C 37	43	79.6	1466	13	BI545222
C 38	42	77.8	185	9	AI608160
C 39	42	77.8	192	13	BI783638
C 40	42	77.8	301	9	AA507304
C 41	42	77.8	319	9	AA596016
C 42	42	77.8	346	9	AA42434
C 43	42	77.8	352	9	AA447045
C 44	42	77.8	369	14	BM716334
C 45	42	77.8	375	12	BF083176

ALIGNMENTS

RESULT 1
AA411980/c
LOCUS zt65g03.sl Soares testis_NHT Homo sapiens cDNA clone IMAGE:727252
DEFINITION 3', mRNA sequence.
ACCESSION AA411980
VERSION AA411980.1 GI:2070632
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gaisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Maier, M., Matlin, J., Moore, B.

TITLE
JOURNAL
COMMENT

, Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2072 Std Error: 0.00
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 306.
Location/Qualifiers

FEATURES
source

1. 318
/organism="Homo sapiens"

/db_xref="GDB:5924164"

/db_xref="taxon:9606"

/clone="IMAGE:727252"

/clone.lib="Soares_testis_NHT"

/sex="male"

/lab_host="DH10B"

BASE COUNT

90 a 72 c 58 g 98 t

ORIGIN

Alignment Scores:

Pred. No.: 246 Length: 318
Score: 47.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 87.04% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-25 (1-10) x AA411980 (1-318)

OY 1 ThrGlyGlyAsnSerGlySerProval 9

Db 219 ACGGAGGAACTCTGCGACTCTATT 193

RESULT 2

LOCUS BH400391/c 426 bp DNA linear GSS 11-DEC-2001

DEFINITION AG-ND-147H4.TF ND-TAM Anopheles gambiae genomic clone AG-ND-147H4,
DNA sequence.

ACCESSION BH400391

VERSION BH400391.1 GI:17346607

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;

Anopheles.

1 (bases 1 to 426)

Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.
Direct Submission of BAC-end sequences from Anopheles gambiae
Unpublished (2001)

Other GSSs: AG-ND-147H4.TR

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@stir.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. 426

/organism="Anopheles gambiae"

/strain="PEST"

/db_xref="taxon:7165"

/clone="AG-ND-147H4"

/clone.lib="ND-TAM"

/note="Vector: pECBAC1, Site_1: HindIII"

BASE COUNT

111 a 79 c 76 g 160 t

ORIGIN

Alignment Scores:

Pred. No.: 330 Length: 426
Score: 47.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 87.04% Indels: 0
DB: 17 Gaps: 0

US-10-008-355-25 (1-10) x BH400391 (1-426)

OY 1 ThrGlyGlyAsnSerGlySerProval 9

Db 81 ACGGAGGAACTCTGCGACTCTCAATT 55

RESULT 3

LOCUS BB622700

DEFINITION BB622700 RIKEN full-length enriched, adult male olfactory brain Mus

musculus cDNA clone 6430525124 5', mRNA sequence.

ACCESSION BB622700.1 GI:16461651

VERSION BB622700

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 608)

Arakawa, T., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hiramoto, K., Horii, F., Ishii, Y., Ito, M., Kawai, J., Konno, H., Kouda, M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M., Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T., Muramatsu, M. and Hayashizaki, Y.

RIKEN Mouse ESTs (Arakawa, T., et al. 2001)

Unpublished (2001)

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@sc.riken.go.jp,

URL: http://genome-gsc.riken.go.jp/

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagl, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubara, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.
e mouse tissues.

FEATURES

SOURCE

Location/Qualifiers
1. 608
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="6430525L24"
/clone_lib="RIKEN full-length enriched, adult male olfactory brain"
/sex="male"
/tissue_type="olfactory brain"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: Salt; Site_2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGCGGCCGACATCGAGTTTCTTTTCTTTTNN 3']. cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTTAATTAAATTCCTCCCTCCCTCC 3']. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 170 a 112 c 145 g 181 t
ORIGIN

Alignment Scores:

Pred. No.: 473 Length: 608
Score: 47.00 Matches: 8
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 80.00% Mismatches: 0
Query Match: 87.04% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-25 (1-10) x BB622700 (1-608)

QY 1 ThrGlyAsnSerGlySerProValPhe 10

Db 561 TCTGTGTAATCTGTGAACCAACCAAGTTT 590

RESULT 4

BF694219/c

LOCUS 611 bp mRNA linear EST 22-DEC-2000

DEFINITION 602082642F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4247003 5',

mRNA sequence.

ACCESSION BF694219

VERSION BF694219.1 GI:11979627
EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Carnivora; Euteleostomi; 1 (bases 1 to 611)
TITLE NIH-MGC <http://imgc.ncbi.nlm.nih.gov/>.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLCM1063 row: d column: 12
High quality sequence stop: 604.

FEATURES

SOURCE

Location/Qualifiers
1. 611
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4247003"
/clone_lib="NIH_MGC_81"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB (Clontech); Site_1: SfiI (ggcgctcgcc); Site_2: SfiI (ggcgctcgcc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCCGAGCGGCCGACATG-dT(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 1.0-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 186 a 144 c 157 g 124 t
ORIGIN

Alignment Scores:

Pred. No.: 692 Length: 611
Score: 46.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 85.19% Indels: 0
DB: 12 Gaps: 0

US-10-008-355-25 (1-10) x BF694219 (1-611)

QY 2 GlyGlyAsnSerGlySerProValPhe 10

Db 458 GGGGTAATCTGTGCGCCAGTTTTC 432

RESULT 5

AO699991

LOCUS 371 bp DNA

DEFINITION HS_5334_A2_G04_SPEE RPCI-11 Human Male BAC Library GSS 06-JUL-1999

genomic clone Plate=910 Col=8 Row=M, DNA sequence.

ACCESSION AO699991

KEYWORDS GI:5390239

SOURCE GSS.

ORGANISM human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Carnivora; Euteleostomi;

1 (bases 1 to 371)

REFERENCE

AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

Sequence-tagged connectors: A sequence approach to mapping and

scanning the human genome
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (<http://inforesgen.com>). BAC end Web Server: <http://www.htsc.washington.edu>
Place: 910 row: M column: 8
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 371.
Location/Qualifiers
1.371
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=910 Col=8 Row=M"
/clone.lib="RPCT-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"
BASE COUNT 87 a 88 c 79 g 117 t
ORIGIN
Alignment Scores:
Pred. No.: 609 Length: 371
Score: 45.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 83.33% Indels: 0
DB: 17 Gaps: 0
US-10-008-355-25 (1-10) x AC699991 (1-371)
QY 1 ThrGlyGlyAsnSerGlySerProval 9
|||||
Db 343 AC6GGGGTAACTCCGCTGCACTGTC 369
RESULT 6
BQ327884 497 bp mRNA linear EST 17-MAY-2002
LOCUS MR4-RT0045-250401-009-g06 RT0045 Homo sapiens cDNA, mRNA sequence.
ACCESSION BQ327884
VERSION BQ327884.1 GI:20945640
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 497)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
JOURNAL MEDLINE 20202663
COMMENT Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=MR4&cl2=MR4-RT0045-250401-009-g06&cl3=2001-04-25&cl4=1>)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 37.
Location/Qualifiers
1.497
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone.lib="RT0045"
/dev-stage="Adult"
/note="Organ: Kidney; tumor: Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196.716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 54 a 164 c 176 g 102 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 818 Length: 497
Score: 45.00 Matches: 8
Percent Similarity: 90.00% Conservative: 1
Best Local Similarity: 80.00% Mismatches: 1
Query Match: 83.33% Indels: 0
DB: 14 Gaps: 0
US-10-008-355-25 (1-10) x BQ327884 (1-497)
QY 1 ThrGlyGlyAsnSerGlySerProvalp 10
|||||
Db 81 AC6GGTGGCGAGACAGTACCTGCTTC 110
RESULT 7
BG714149 796 bp mRNA linear EST 08-MAY-2001
LOCUS 602674594F1 NIH_MGC_96 Homo sapiens cDNA clone IMAGE:4797094 5', mRNA sequence.
ACCESSION BG714149
VERSION BG714149.1 GI:13993080
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 796)
NIH-MGC <http://mgc.nci.nih.gov/>.
Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-rt@mail.nih.gov
Tissue Procurement: Miklos Palakovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiroki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLUMINA)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILLUMINA at: <http://image.llnl.gov>
Plate: LLM10682 row: 1 column: 23
High quality sequence stop: 677.
Location/Qualifiers
1.796
source

```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4797094"
/clone_lib="NIH_MGC_96"
/tissue_type="hypothalamus"
/lab_host="DH10B"
/notes="Organ: Brain; Vector: pBluescript (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTN-3',
size-selected for average insert size 2.3 kb and
normalized to ROF 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/MHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT      172 a      226 c      275 g      123 t

ORIGIN
Alignment Scores:
Pred. No.:      1.32e+03      Length:      796
Score:          45.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 88.89%      Mismatches: 0
Query Match:    83.33%      Indels:      0
DB:             12      Gaps:      0

US-10-008-355-25 (1-10) x BG174149 (1-796)

QY      1 ThGlyGlyAnsSerGlySerProval 9
|||||
Db      770 ACAGGTGGCAACAGCGTGCACCTGTG 744

RESULT 8
BF128636      914 bp      mRNA      linear      EST 24-OCT-2000
LOCUS      601810895F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4054078 5',
DEFINITION      mRNA sequence.
ACCESSION      BF128636
VERSION      BF128636.1 GI:10967676
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH_MGC http://mgc.nci.nih.gov/
1 (bases 1 to 914)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Louis M. Straudt, M.D., Ph.D.
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM893 row: m column: 23
High quality sequence stop: 647.
Location/Qualifiers
1. 914

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4054078"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAAGGAG(g). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
```

```
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT      257 a      237 c      266 g      154 t

ORIGIN
Alignment Scores:
Pred. No.:      1.51e+03      Length:      914
Score:          45.00      Matches:      8
Percent Similarity: 90.00%      Conservative: 1
Best Local Similarity: 80.00%      Mismatches: 1
Query Match:    83.33%      Indels:      0
DB:             12      Gaps:      0

US-10-008-355-25 (1-10) x BF128636 (1-914)

QY      1 ThGlyGlyAnsSerGlySerProvalPhe 10
|||||
Db      701 ACAGCGCGCAATTCGGCAATCCTCTTTC 730

RESULT 9
BG115436      1108 bp      mRNA      linear      EST 30-JAN-2001
LOCUS      602316560F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4416831 5',
DEFINITION      mRNA sequence.
ACCESSION      BG115436
VERSION      BG115436.1 GI:12608942
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NIH_MGC http://mgc.nci.nih.gov/
1 (bases 1 to 1108)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM10148 row: h column: 16
High quality sequence stop: 680.
Location/Qualifiers
1. 1108

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4416831"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: small intestine; Vector: pCW-SPOPT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dT primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT      203 a      367 c      312 g      226 t

ORIGIN
Alignment Scores:
Pred. No.:      1.84e+03      Length:      1108
Score:          45.00      Matches:      7
Percent Similarity: 100.00%      Conservative: 2
Best Local Similarity: 77.78%      Mismatches: 0
Query Match:    83.33%      Indels:      0
DB:             12      Gaps:      0

US-10-008-355-25 (1-10) x BG115436 (1-1108)
```

OY 2 G1yG1yAsnserglyserProvalphe 10
|||||
Db 615 GGGGGGACACGACCAATCCGATTTC 641

RESULT 10
LOCUS BF893687/c 237 bp mRNA linear EST 18-JAN-2001
DEFINITION OVI-MT0166-131100-476-b02 MT0166 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF893687
VERSION BF893687.1 GI:12285146
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo.
1 (bases 1 to 237)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV1-MT0166-131100-476-b02&t3=2000-11-13&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 182.
Location/Qualifiers
1..237
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MT0166"
/dev_stage="Adult"
/note="Organ: marrow; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 40 a 60 c 52 g 85 t
ORIGIN

Alignment Scores:
Pred. No.: 565 Length: 237
Score: 44.00 Matches: 7
Percent Similarity: 90.00% Conservative: 2
Best Local Similarity: 70.00% Mismatches: 1
Query Match: 81.48% Indels: 0
DB: 12 Gaps: 0
US-10-008-355-25 (1-10) x BF893687 (1-237)
OY 1 ThrG1yG1yAsnserglyserProvalphe 10
|||||
Db 236 ACAGGGGGAACACTGGAAGCCACTTTT 207

RESULT 11
LOCUS AW060086/c 279 bp mRNA linear EST 30-MAR-2000

DEFINITION 687001G07.x1 687 - Early embryo from Delaware Zea mays cDNA, mRNA sequence.
ACCESSION AW060086
VERSION AW060086.1 GI:5936818
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: PACC clade: Panicoideae: Andropogoneae: Zea.
1 (bases 1 to 279)
REFERENCE Walbot,V.
Maize ESTs from various cDNA libraries sequenced at Stanford University
Unpublished (1999)
JOURNAL
TITLE Walbot V
CONTACT: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 687001 row: G column: 07.
Location/Qualifiers
1..279
/organism="Zea mays"
/cultivar="Illinois High Oil"
/db_xref="taxon:4577"
/clone_lib="687 - Early embryo from Delaware"
/issue_type="embryo"
/dev_stage="14, 21, 28, and 35 days after pollination"
/lab_host="E. coli SOLR"
/note="Organ: embryo; Vector: plusescript SK; Site.1: XhoI ; Site.2: EcoRI; library was prepared by stragene using the uni-ZAP XR system (Stratagene BN937328-12). Clones were picked by a Q-bot after blue/white selection (ampicillin resistance - use 100 micograms/microliter). Developed from a pool of equal amounts of RNA from developing embryos sampled at 14, 21, 28 and 35 days after pollination of the Illinois High Oil Maize Strain cycle 90. This closed strain has been selected for high oil concentration for 90 generations and originates from the 1890s era open pollinated variety Burr's White"
BASE COUNT 61 a 78 c 68 g 72 t
ORIGIN

Alignment Scores:
Pred. No.: 666 Length: 279
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 81.48% Indels: 0
DB: 10 Gaps: 0
US-10-008-355-25 (1-10) x AW060086 (1-279)
OY 1 ThrG1yG1yAsnserglyserPro 8
|||||
Db 155 ACAGAGAGAAACAGTGCGACGCCG 132

RESULT 12
LOCUS BJ099916/c 298 bp mRNA linear EST 12-DEC-2001
DEFINITION BJ099916 N1BB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone X115108.3', mRNA sequence.
ACCESSION BJ099916
VERSION BJ099916.1 GI:17602460
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Amphibia: Batrachia: Anura: Mesobatrachia: Pipidoidea: Pipidae: Xenopodinae: Xenopus.
298 bp mRNA linear EST 12-DEC-2001

REFERENCE 1 (bases 1 to 298)
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@gene.nig.ac.jp.

FEATURES
Source
Location/Qualifiers
1..298
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="X15108"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/note="Vector: pBSR3; Site.1: NotI; Site.2: EcoRI; cDNAs were oligo-dT primed and directionally cloned. Staging and according to Nieuwkoop and Faber. Library is subtracted and was constructed by N. Garrett and A.M. Zorn, (Wellcome/CRC Institute)."

BASE COUNT 102 a 72 c 62 g 58 t 4 others
ORIGIN

Alignment Scores:
Pred. No.: 712 Length: 298
Score: 44.00 Matches: 8
Percent Similarity: 80.00% Conservative: 0
Best Local Similarity: 80.00% Mismatches: 2
Query Match: 81.48% Indels: 0
Gaps: 0

US-10-008-355-25 (1-10) x BU099916 (1-298)

QY 1 ThrGlyGlyAsnSerGlySerProValPhe 10
|||||
Db 35 ACTGGGGCAATAGTGGGCCCTTTT 6

RESULT 13
LOCUS AA476126 403 bp mRNA linear EST 18-JUN-1997
DEFINITION vhl7d08.t1 Soares_mammary_gland_NBMG Mus musculus cDNA clone
IMAGE:875127 5', mRNA sequence.
ACCESSION AA476126
VERSION AA476126.1 GI:2203977
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 403)
Marrera,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucab,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 850L, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:515207

```

FEATURES                               Seq primer: -28ml3 rev2 ET from Amersham
                                         High quality sequence stop: 382.
                                         Location/Qualifiers
                                         1..403
                                           /organism="Mus musculus"
                                           /strain="C57BL/6J"
                                           /db_xref="taxon:10090"
                                           /clone="IMAGE:875727"
                                           /clone_id="Soares_mammary_gland_NbMWG"
                                           /sex="male"
                                           /tissue_type="mammary gland"
                                           /dev_stage="4 weeks"
                                           /lab_host="DH10B"
                                           /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                                           ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                                           RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                                           primer [5',
                                           TGTTACCAATCTGAAGCGGGCCGCCAATGTCTTTTTTTTTTTTTTTTTTTT
                                           T 3']; double-stranded cDNA was ligated to Eco RI
                                           adaptors (Pharmacia), digested with Not I and cloned into
                                           the Not I and Eco RI sites of the modified pT7T3 vector.
                                           RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                                           constructed and normalized by Bento Soares and M.Fatima
                                           Bonaldo."
BASE COUNT      115 a          61 c          86 g          141 t
ORIGIN
Alignment Scores:
Pairing No.:           965              Length:             403
Score:                44.00            Matches:              8
Percent Similarity:   100.00%          Conservative:         1
Best Local Similarity: 88.89%          Mismatches:           0
Query Match:          81.48%           Indels:               0
DB:                    9                Gaps:                 0

US-10-008-355-25 (1-10) x AA476126 (1-403)
OY    2 Glycylasnserylseryprovalphe 10
      |||:::|||||||
DB    84 GGTGGCTCTTCGATCCCATATT 110

RESULT 14
W72307                                420 bp      mRNA       linear     EST 16-OCT-1996
LOCUS      w72307
DEFINITION IMAGE:z60q08.s1 Soares,fetal heart,NbHn19M Homo sapiens cDNA clone
VERSION    W72307
KEYWORDS   W72307.1 GI:1382930
SOURCE     EST.
ORGANISM   human.
            Homo saplens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 420)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevasakis,E., Waterston
            ,R., Williamson,A., Wohlmann,P. and Wilson,R.
            The WashU-Merck EST Project
TITLE      Unpublished (1995)
JOURNAL    Contact: Wilson RK
COMMENT    Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert length: 389 Std Error: 0.00
            Seq primer: mob.RECA+ET
            High quality sequence stop: 325.
            Location/Qualifiers
            1..420

```

```
/organism="Homo sapiens"
/db_xref="CDB:1270461"
/db_xref="taxon:9606"
/clone="IMAGE:345086"
/clone_lib="Soares_fetal_heart_NbH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCCGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Falima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbH19W."
BASE COUNT      129 a      121 c      88 g      79 t      3 others
ORIGIN
Alignment Scores:
Pred. No.:      1.01e+03      Length:      420
Score:          44.00      Matches:      8
Percent Similarity: 88.89%      Conservative: 0
Best Local Similarity: 88.89%      Mismatches:  1
Query Match:    81.48%      Indels:      0
DB:             14      Gaps:          0
US-10-008-355-25 (1-10) x W2307 (1-420)
Oy      1      ThrGlyAsnSerGlySerProval 9
|||||:|||||:|||||:|||||:|||||
Db      201      ACAGGTGCACACGCGANTCCCGTG 227
RESULT 15
AW743569      432 bp      mRNA      linear      EST 27-APR-2000
LOCUS      ur23d03.y1 Soares_mouse_MMBP Mus musculus cDNA clone IMAGE:3025157
DEFINITION      5', mRNA sequence.
ACCESSION      AW743569
VERSION      AW743569.1 GI:7655383
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 432)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CCAP).
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1053865
Seq primer: -40RP from Gibco
High quality sequence stop: 404.
Location/Qualifiers
1. 432
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3025157"
/clone_lib="Soares_mouse_MMBP"
/lab_host="DH10B (phage-resistant)"
/note="Organ: bone (pooled); Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: NotI; Site_2: EcoRI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTACCAATCTGAAGTGGAGCGCCGATCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
```

```
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M.Falima
Bonaldo."
BASE COUNT      125 a      77 c      96 g      134 t
ORIGIN
Alignment Scores:
Pred. No.:      1.04e+03      Length:      432
Score:          44.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 1
Best Local Similarity: 88.89%      Mismatches:  0
Query Match:    81.48%      Indels:      0
DB:             10      Gaps:          0
US-10-008-355-25 (1-10) x AW743569 (1-432)
Oy      2      GlyGlyAsnSerGlySerProvalphe 10
|||||:|||||:|||||:|||||:|||||
Db      352      GGTGGCTCTTCGTGATCCCGATATTT 378
Search completed: December 20, 2002, 15:48:51
Job time : 1535.05 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 05:28:30 : Search time 49.4737 Seconds
(without alignments)
26.934 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TGGNSGSPVF 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	10	23	AAO15228 Porphyromonas ging
2	54	100.0	52	23	AAO15206 Porphyromonas ging
3	54	100.0	52	23	AAO15207 Staphylococcus aur
4	54	100.0	213	17	AAO15205 V8 mature protease
5	54	100.0	214	17	AAO15205 V8 mature protease
6	54	100.0	215	17	AAO15205 V8 mature protease
7	54	100.0	336	13	AAO15205 V8 mature protease
8	54	100.0	344	17	AAO15205 V8 mature protease
9	54	100.0	344	18	AAO15205 V8 mature protease
10	54	100.0	357	13	AAO15205 V8 mature protease

11	54	100.0	392	17	AAO15205 V8 mature protease
12	54	100.0	532	17	AAO15205 V8 mature protease
13	54	100.0	532	18	AAO15205 V8 mature protease
14	54	100.0	537	18	AAO15205 V8 mature protease
15	54	100.0	712	23	AAO15205 V8 mature protease
16	49	90.7	282	22	AAO15205 V8 mature protease
17	49	90.7	311	23	AAO15205 V8 mature protease
18	48	88.9	9	23	AAO15205 V8 mature protease
19	44	81.5	252	20	AAO15205 V8 mature protease
20	44	81.5	252	23	AAO15205 V8 mature protease
21	44	81.5	284	20	AAO15205 V8 mature protease
22	44	81.5	284	23	AAO15205 V8 mature protease
23	44	81.5	336	20	AAO15205 V8 mature protease
24	43	79.6	221	22	AAO15205 V8 mature protease
25	43	79.6	222	12	AAO15205 V8 mature protease
26	43	79.6	222	13	AAO15205 V8 mature protease
27	43	79.6	222	22	AAO15205 V8 mature protease
28	43	79.6	222	22	AAO15205 V8 mature protease
29	43	79.6	222	22	AAO15205 V8 mature protease
30	43	79.6	222	22	AAO15205 V8 mature protease
31	43	79.6	222	22	AAO15205 V8 mature protease
32	43	79.6	222	22	AAO15205 V8 mature protease
33	43	79.6	222	22	AAO15205 V8 mature protease
34	43	79.6	222	22	AAO15205 V8 mature protease
35	43	79.6	222	22	AAO15205 V8 mature protease
36	43	79.6	222	22	AAO15205 V8 mature protease
37	43	79.6	222	22	AAO15205 V8 mature protease
38	43	79.6	222	22	AAO15205 V8 mature protease
39	43	79.6	222	22	AAO15205 V8 mature protease
40	43	79.6	222	22	AAO15205 V8 mature protease
41	43	79.6	222	22	AAO15205 V8 mature protease
42	43	79.6	222	22	AAO15205 V8 mature protease
43	43	79.6	222	22	AAO15205 V8 mature protease
44	43	79.6	222	22	AAO15205 V8 mature protease
45	43	79.6	222	22	AAO15205 V8 mature protease

ALIGNMENTS

RESULT 1
AAO15228
ID AAO15228 standard; Peptide: 10 AA.
XX
AC AAO15228;
XX
DT 05-SEP-2002 (first entry)
XX
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) active site 1.
XX
KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage; active site;
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;
KW periodontitis.
XX
OS Porphyromonas gingivalis.
XX
PN WO200238742-A2.
XX
PD 16-MAY-2002.
XX
PF 08-NOV-2001; 2001WO-0546782.
XX
PR 08-NOV-2000; 2000US-246827P.
XX
(UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Travis J, Potempa JS, Bandula A, Bugno M;
XX
DR WPI; 2002-490075/52.
XX
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT the dipeptidylpeptidase for protecting an animal from periodontal
PT disease caused by Porphyromonas gingivalis

XX Claim 5; Page 32; 65pp; English.
PS
XX
CC The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present amino acid sequence represents an active
CC site region of the Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7)
CC enzyme.
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 54; DB 23; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.031;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCNCGSPVF 10
DB 1 TGCNCGSPVF 10
RESULT 2
AAOI5206
ID AAOI5206 standard; Protein; 52 AA.
XX
AC AAOI5206;
XX
DT 05-SEP-2002 (first entry)
XX
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.
XX
XX Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;
KM periodontitis.
XX
OS Porphyromonas gingivalis.
XX
PN WO200238742-A2.
XX
PD 16-MAY-2002.
XX
PF 08-NOV-2001; 2001WO-US46782.
XX
PR 08-NOV-2000; 2000US-246827P.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Travis J, Potempa JS, Banbula A, Bugno M;
XX
DR WPI; 2002-490075/52.
XX
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT the dipeptidylpeptidase for protecting an animal from periodontal
PT disease caused by Porphyromonas gingivalis -
XX
XX Example 6; Fig 5; 65pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a

CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present amino acid sequence represents the
CC C-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7
CC (DPP-7) enzyme.
XX
SQ Sequence 52 AA;
Query Match 100.0%; Score 54; DB 23; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCNCGSPVF 10
DB 1 TGCNCGSPVF 10
RESULT 3
AAOI5207
ID AAOI5207 standard; Protein; 52 AA.
XX
AC AAOI5207;
XX
DT 05-SEP-2002 (first entry)
XX
DE Staphylococcus aureus V8 endopeptidase C-terminal region.
XX
XX Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;
KM periodontitis; V8 endopeptidase.
XX
OS Staphylococcus aureus.
XX
PN WO200238742-A2.
XX
PD 16-MAY-2002.
XX
PF 08-NOV-2001; 2001WO-US46782.
XX
PR 08-NOV-2000; 2000US-246827P.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Travis J, Potempa JS, Banbula A, Bugno M;
XX
DR WPI; 2002-490075/52.
XX
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT the dipeptidylpeptidase for protecting an animal from periodontal
PT disease caused by Porphyromonas gingivalis -
XX
XX Example 6; Fig 5; 65pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present amino acid sequence represents the
CC C-terminal region of the Staphylococcus aureus V8 endopeptidase enzyme.
XX
SQ Sequence 52 AA;
Query Match 100.0%; Score 54; DB 23; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGCNCGSPVF 10

Db 1 TCGNSGSPVF 10

RESULT 4
ID AAR91042 standard; Peptide: 213 AA.
XX AAR91042;
XX
DT 23-MAY-1996 (first entry)
XX
DE V8 mature protease (aal-213).
XX
XX Linker peptide; V8 protease; Staphylococcus aureus;
KW recombinant protein; fusion protein; beta-galactosidase;
KW Escherichia coli; transposon Tn903;
KW aminoglycoside 3'-phosphotransferase.
XX
OS Staphylococcus aureus strain V8 (ATCC 27733).
XX EP700995-A2.
XX
PD 13-MAR-1996.
XX
XX 06-SEP-1995; 95EP-0306235.
XX
XX 07-NOV-1994; 94JP-0296028.
PR 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR) SUNTORY LTD.
PA
XX Ohsuye K, Yabuta M;
PI
XX WPI; 1996-141021/15.
DR
XX
XX Prod. of recombinant polypeptide(s) - using host cells transformed
PT with a gene coding for the desired polypeptide fused to a protective
PT polypeptide
XX
XX Example 8; Fig 14a; 44pp; English.
PS
XX
XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC and C are protective polypeptides, B is Staphylococcus aureus mature
CC V8 protease lacking the C-terminal repeat region, and L is a linker
CC (AAR91032). It was produced as an inclusion body in Escherichia coli
CC host cells, and was cleaved using the E. coli OmpT protease to
CC yield active V8 protease. Extension of the C-terminal end of the
CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)
CC i.e. up to residue Phe-215, also resulted in the formation of
CC inclusion bodies in E. coli. Any further extension gave a soluble
CC product which exhibited protease activity that repressed growth of
CC the host cells.
XX
SQ Sequence 213 AA;

Query Match 100.0%; Score 54; DB 17; Length 213;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 10
DB 165 TCGNSGSPVF 174

RESULT 5
ID AAR91043 standard; Peptide: 214 AA.
XX AAR91043;
XX
DT 23-MAY-1996 (first entry)
XX
DE V8 mature protease (aal-214).

XX
KW Linker peptide; V8 protease; Staphylococcus aureus;
KW recombinant protein; fusion protein; beta-galactosidase;
KW Escherichia coli; transposon Tn903;
KW aminoglycoside 3'-phosphotransferase.
XX
OS Staphylococcus aureus strain V8 (ATCC 27733).
XX EP700995-A2.
XX
PD 13-MAR-1996.
XX
XX 06-SEP-1995; 95EP-0306235.
XX
XX 07-NOV-1994; 94JP-0296028.
PR 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR) SUNTORY LTD.
PA
XX Ohsuye K, Yabuta M;
PI
XX WPI; 1996-141021/15.
DR
XX
XX Prod. of recombinant polypeptide(s) - using host cells transformed
PT with a gene coding for the desired polypeptide fused to a protective
PT polypeptide
XX
XX Example 8; Fig 14b; 44pp; English.
PS
XX
XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC and C are protective polypeptides, B is Staphylococcus aureus mature
CC V8 protease lacking the C-terminal repeat region, and L is a linker
CC (AAR91032). It was produced as an inclusion body in Escherichia coli
CC host cells, and was cleaved using the E. coli OmpT protease to
CC yield active V8 protease. Extension of the C-terminal end of the
CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)
CC i.e. up to residue Phe-215, also resulted in the formation of
CC inclusion bodies in E. coli. Any further extension gave a soluble
CC product which exhibited protease activity that repressed growth of
CC the host cells.
XX
SQ Sequence 214 AA;

Query Match 100.0%; Score 54; DB 17; Length 214;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPVF 10
DB 165 TCGNSGSPVF 174

RESULT 6
ID AAR91044 standard; Peptide: 215 AA.
XX AAR91044;
XX
DT 23-MAY-1996 (first entry)
XX
DE V8 mature protease (aal-215).
XX
XX Linker peptide; V8 protease; Staphylococcus aureus;
KW recombinant protein; fusion protein; beta-galactosidase;
KW Escherichia coli; transposon Tn903;
KW aminoglycoside 3'-phosphotransferase.
XX
OS Staphylococcus aureus strain V8 (ATCC 27733).
XX EP700995-A2.
XX
PD 13-MAR-1996.
XX

PF 06-SEP-1995; 95EP-0306235.
XX
PR 07-NOV-1994; 94JP-0296028.
XX 07-SEP-1994; 94JP-0238595.
XX
PA (SUNR) SUNTORY LTD.
XX
PI Ohsuye K, Yabuta M;
XX
DR WPI; 1996-141021/15.
XX
XX Prod. of recombinant polypeptide(s) - using host cells transformed
PT with a gene coding for the desired polypeptide fused to a protective
PT polypeptide
XX
PS Example 8; Fig 14C; 44pp; English.
XX
CC A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC and C are protective polypeptides, B is Staphylococcus aureus mature
CC V8 protease lacking the C-terminal repeat region, and L is a linker
CC (AAR91032). It was produced as an inclusion body in Escherichia coli
CC host cells, and was cleaved using the E. coli OmpT protease to the
CC yield active V8 protease. Extension of the C-terminal end of the
CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)
CC i.e. up to residue phe-215, also resulted in the formation of
CC inclusion bodies in E. coli. Any further extension gave a soluble
CC product which exhibited protease activity that repressed growth of
CC the host cells.
XX
SQ Sequence 215 AA;
XX
Query Match 100.0%; Score 54; DB 17; Length 215;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGCNCGSPVF 10
XXXXXXXXXXXX
DB 165 TGCNCGSPVF 174
XX
RESULT 7
AAR29644
ID AAR29644 standard; Protein; 336 AA.
XX
AC AAR29644;
XX
DT 11-FEB-1993 (first entry)
XX
DE Protease from S. Aureus.
XX
DE Protease: PCR; amplify; staphylococcus.
XX
KM Protease; PCR; amplify; staphylococcus.
XX
OS Staphylococcus aureus.
XX
XX Key Location/Qualifiers
FH Peptide 1..68
FT /label= Signal_peptide
FT Protein 69..336
FT /label= Protease
XX
XX JP04211370-A.
XX
PD 03-AUG-1992.
XX
PF 19-FEB-1991; 91JP-0024633.
XX
PR 20-FEB-1990; 90JP-0040398.
XX
PA (SHIO) SHIONOGI & CO LTD.
XX
DR WPI; 1992-304938/37.
XX
DR N-PSDB; AAQ27988.
XX

PT Novel protease prep'd. using Bacillus or Saccharomyces host -
PT capable of cleaving peptide bond at carboxyl terminus of glutamic
PT acid residues in polypeptide(s)
XX
XX Disclosure; Page 15-16; 25pp; Japanese.
XX
XX The sequences given in AAR26842 and AAR29644 are proteases which were
CC isolated from Staphylococcus aureus strains. The DNA sequences
CC encoding these proteins were isolated by PCR using the primer
CC sequences given in AAQ27960-86. The protease specifically cleaves
CC the peptide bond at the C-terminus of the glutamic acid residue in
CC polypeptide.
XX
SQ Sequence 336 AA;
XX
Query Match 100.0%; Score 54; DB 13; Length 336;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGCNCGSPVF 10
XXXXXXXXXXXX
DB 233 TGCNCGSPVF 242
XX
RESULT 8
AAR91033
ID AAR91033 standard; Protein; 344 AA.
XX
AC AAR91033;
XX
DT 23-MAY-1996 (first entry)
XX
DE Beta-galactosidase-V8 protease fusion protein.
XX
KW V8 protease: Staphylococcus aureus; recombinant protein;
KW fusion protein; beta-galactosidase; Escherichia coli.
XX
OS Chimeric Escherichia coli;
OS Chimeric Staphylococcus aureus strain V8 (ATCC 27733).
XX
FH Key Location/Qualifiers
FT Region 1..124
FT /note= "beta-galactosidase region"
FT Region 125..344
FT /note= "mature V8 protease without the repeat
FT region"
XX
PN EP700995-A2.
XX
PD 13-MAR-1996.
XX
PE 06-SEP-1995; 95EP-0306235.
XX
PR 07-NOV-1994; 94JP-0296028.
PR 07-SEP-1994; 94JP-0238595.
XX
PA (SUNR) SUNTORY LTD.
XX
PI Ohsuye K, Yabuta M;
XX
DR WPI; 1996-141021/15.
XX
XX Prod. of recombinant polypeptide(s) - using host cells transformed
PT with a gene coding for the desired polypeptide fused to a protective
PT polypeptide
XX
PS Example 2; Page 12-13; 44pp; English.
XX
CC Fusion proteins (AAR91033 and AAR91034) were constructed comprising
CC an Escherichia coli beta-galactosidase derivative (protective
CC polypeptide) fused at its C-terminal end to the S. aureus mature
CC V8 protease without or with the repeat region. The constructs
CC were inserted into vector pc97S4DnCT(G)R6, yielding pV8RPT(-) and

CC pV8RPT(+), respectively. Both constructs yielded active protease
CC when expressed in E. coli JM101 transformants.

XX Sequence 344 AA;

Query Match 100.0%; Score 54; DB 17; Length 344;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10

DB 289 TCGNSGSPVF 298

RESULT 9

AAM22218

ID AAM22218 standard; Protein: 344 AA.

AC AAM22218;

DT 11-SEP-1997 (first entry)

DE Protein encoded by pV8RPT(-) construct.

XX Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;

KW truncation; wild type; PCR; polymerase chain reaction; amplification;

KM proteolytic activity; fusion protein; beta-galactosidase; urea.

XX Chimeric - Escherichia coli.

OS Chimeric - Staphylococcus aureus.

FH Key Location/Qualifiers

FT Region 1..100 /note="E. coli beta-galactosidase portion"

FT Region 101..120

FT /note="synthetic R6 linker"

FT Region 125..344 /note="truncated S. aureus V8 protease portion"

FT EP745669-A2.

PN 04-DEC-1996.

PD 31-MAY-1996; 96EP-0303939.

PF 02-JUN-1995; 95JP-0170086.

PR (SUNR) SUNTORY LTD.

PA Ohshye K, Yabuta M;

PI WPI: 1997-013693/02.

DR Staphylococcus aureus V8 protease mutants - with increased

PT resistance to denaturation

PS Claim 2; Page 13-14; 42pp; English.

XX The invention relates to new mutant Staphylococcus aureus V8 proteases

CC which have enzyme activity even under environmental conditions which

CC promote protein denaturation. The mutants are based on 3 truncated V8

CC proteases lacking 48 (AAM22218), 56 (AAM2219) or 53 (AAM2220) amino

CC acids from the C-terminal of the wild type protease. The mutants also

CC contain amino acid substitutions, especially D44E, N71S and/or R147K.

CC The protein sequence shown here represents a chimeric protein

CC comprising a truncated Staphylococcus aureus V8 protease lacking the

CC prepro and C-terminal 48 amino acids linked, via a synthetic linker,

CC downstream of the E. coli beta-galactosidase. The S. aureus portion of

CC the chimaera was amplified by the primers AAT73254-5 from wild type

CC sequence. The coding sequence was then used to generate mutants of the

CC V8 protease which retain their levels of activity in the presence of a

CC higher concentration of protein denaturant e.g. 5 M urea.

SQ Sequence 344 AA;

Query Match 100.0%; Score 54; DB 18; Length 344;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10

DB 289 TCGNSGSPVF 298

RESULT 10

AAR26842

ID AAR26842 standard; Protein: 357 AA.

AC AAR26842;

DT 11-FEB-1993 (first entry)

DE Protease from S. Aureus ATCC12600.

XX Protease; PCR; amplify; Staphylococcus.

OS Staphylococcus aureus.

FH Key Location/Qualifiers

FT Peptide 1..68 /label= Signal_peptide

FT Protein 69..358 /label= Protease

PN JP04211370-A.

PD 03-AUG-1992.

PF 19-FEB-1991; 91JP-0024633.

PR 20-FEB-1990; 90JP-0040398.

PA (SHITO) SHIONOGI & CO LTD.

PI WPI: 1992-304938/37.

DR N-PSDB; AAQ27987.

PT Novel protease prep'd. using Bacillus or Saccharomyces host -

PT capable of cleaving peptide bond at carboxyl terminus of glutamic

PT acid residues in polypeptide(s)

PS Disclosure: Page 13-15; 25pp; Japanese.

CC The sequences given in AAR26842 and AAR29644 are proteases which were

CC isolated from Staphylococcus aureus strains. The DNA sequences

CC encoding these proteins were isolated by PCR using the primer

CC sequences given in AAQ27960-86. The protease specifically cleaves

CC the peptide bond at the C-terminus of the glutamic acid residue in

CC polypeptide.

SQ Sequence 357 AA;

Query Match 100.0%; Score 54; DB 13; Length 357;

Best Local Similarity 100.0%; Pred. No. 1.2;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10

DB 233 TCGNSGSPVF 242

RESULT 11

ID AAR91034 standard; Protein: 392 AA.

AC AAR91034;

```

XX 23-MAY-1996 (first entry)
DT Beta-galactosidase-V8 protease fusion protein.
DE
XX
KM V8 protease; Staphylococcus aureus; recombinant protein;
KM fusion protein; beta-galactosidase; Escherichia coli.
XX
OS Chimeric Escherichia coli;.
XX Chimeric Staphylococcus aureus strain V8 (ATCC 27733).
XX
FH Key
FH Region 1..124
FT /note="beta-galactosidase region"
FT Region 125..392
FT /note="mature V8 protease including the repeat
FT region"
XX
FN EP700995-A2.
XX
PD 13-MAR-1996.
XX
PF 06-SEP-1995; 95EP-0306235.
XX
PR 07-NOV-1994; 94JP-0296028.
PR 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR ) SUNTORY LTD.
XX
PI Ohshuye K, Yabuta M;
XX
DR WPI; 1996-141021/15.
XX
XX Prodn. of recombinant polypeptide(s) - using host cells transformed
PT with a gene coding for the desired polypeptide fused to a protective
PT polypeptide
XX
XX Example 2; Page 14-15; 44pp; English.
XX
PS Fusion proteins (AAR91033 and AAR91034) were constructed comprising
CC an Escherichia coli beta-galactosidase derivative (protective
CC polypeptide) fused at its C-terminal end to the S. aureus mature
CC V8 protease without or with the repeat region. The constructs
CC were inserted into vector pG97S4DhCT(G)R6, yielding pV8RPT(-) and
CC pV8RPT(+), respectively. Both constructs yielded active protease
CC when expressed in E. coli JM101 transformants.
XX
SQ Sequence 392 AA:
QY 1 TGGNSGSPVF 10
DB 289 TGGNSGSPVF 298

```

RESULT 12

AAR91035 standard; Protein: 532 AA.

AAR91035;

23-MAY-1996 (first entry)

Recombinant V8 protease V8D fusion protein.

Linker peptide; V8 protease; Staphylococcus aureus;

recombinant protein; fusion protein; beta-galactosidase;

Escherichia coli; transposon Tn903;

aminoglycoside 3'-phosphotransferase.

```

OS Chimeric Escherichia coli;
OS Chimeric synthetic;
OS Chimeric Staphylococcus aureus strain V8 (ATCC 27733);
OS Chimeric transposon Tn903.
XX
FH Key
FH Region 1..100
FT /note="beta-galactosidase region"
FT Region 101..120
FT /note="R6 linker"
FT Cleavage-site 104..105
FT /note="cleavage site for OmpT protease"
FT Region 125..335
FT /note="V8 protease region"
FT Region 336..356
FT /note="R6 linker"
FT Cleavage-site 339..340
FT /note="cleavage site for OmpT protease"
FT Region 307..532
FT /note="aminoglycoside 3'phosphotransferase
FT region"
XX
FN EP700995-A2.
XX
PD 13-MAR-1996.
XX
PF 06-SEP-1995; 95EP-0306235.
XX
PR 07-NOV-1994; 94JP-0296028.
PR 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR ) SUNTORY LTD.
XX
PI Ohshuye K, Yabuta M;
XX
DR WPI; 1996-141021/15.
XX
XX Prodn. of recombinant polypeptide(s) - using host cells transformed
PT with a gene coding for the desired polypeptide fused to a protective
PT polypeptide
XX
XX Example 3; Page 16-18; 44pp; English.
XX
PS A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC and C are protective polypeptides (derived from Escherichia coli
CC beta-galactosidase and and Tn903 aminoglycoside 3'-phosphotransferase,
CC respectively). B is Staphylococcus aureus mature V8 protease
CC lacking the C-terminal repeat region, and L is a linker peptide
CC (AAR91032). The fusion protein is expressed in inactive form in E.
CC coli. It is then recovered, solubilised and cleaved at the linker
CC peptide regions with a protease intrinsic to the host cells, i.e.
CC OmpT protease, to allow recovery of V8 protease.
XX
SQ Sequence 532 AA:
QY 1 TGGNSGSPVF 10
DB 289 TGGNSGSPVF 298

```

RESULT 13

AAW22219 standard; Protein: 532 AA.

AAW22219;

11-SEP-1997 (first entry)

Protein encoded by pV8D construct.

XX	Mutant: Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;
XX	truncation; wild type; PCR; polymerase chain reaction; amplification;
KW	proteolytic activity; fusion protein; beta-galactosidase; urea.
XX	
OS	Chimeric - Escherichia coli.
OS	Chimeric - Staphylococcus aureus.
XX	
EH	Key
FT	Region
FT	Location/Qualifiers
FT	1..100
FT	/note= "E. coli beta-galactosidase portion"
FT	101..124
FT	/note= "R6 linker sequence"
FT	125..336
FT	/note= "truncated S. aureus V8 protease portion"
FT	337..360
FT	/note= "R6 linker sequence"
FT	361..532
FT	/note= "aminoglucoside 3'-phosphotransferase portion"
XX	
PN	EP745669-A2.
PD	04-DEC-1996.
XX	
PE	31-MAY-1996; 96EP-0303939.
XX	
PR	02-JUN-1995; 95JP-0170086.
PA	(SUNR) SUNTORY LTD.
XX	
PI	Ohshye K, Yabuta M;
DR	WPI; 1997-013693/02.
XX	
PT	Staphylococcus aureus V8 protease mutants - with increased
PT	resistance to denaturation
XX	
PS	Claim 7; Page 16-17; 42pp; English.
XX	
CC	The invention relates to new mutant Staphylococcus aureus V8 proteases
CC	which have enzyme activity even under environmental conditions which
CC	promote protein denaturation. The mutants are based on 3 truncated V8
CC	proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
CC	acids from the C-terminal of the wild type protease. The mutants also
CC	contain amino acid substitutions, especially D44E, N71S and/or R147K.
CC	The protein sequence shown here represents a chimaeric protein
CC	comprising a truncated Staphylococcus aureus V8 protease lacking the
CC	propro and C-terminal 56 amino acids linked, via a synthetic R6 linker,
CC	downstream of the E. coli beta-galactosidase. Also included downstream
CC	of the V8 protease fragment is a second R6 linker and a fragment of the
CC	aminoglucoside 3'-phosphotransferase protein. The chimaeric sequence
CC	was generated by restriction digestion and ligation from the V8RPT(-)
CC	sequence (see AAW22218) by using a natural EcoRV site which removed a
CC	further 8 amino acid from the C-terminus. This truncated V8 protease,
CC	designated VBD, retains its level of activity in the presence of a
CC	higher concentration of protein denaturant e.g. 5 M urea.
XX	
XX	
SO	Sequence 532 AA;
XX	
OY	1 TGGNGSPVF 10
DB	289 TGGNGSPVF 298
XX	
XX	100.0%; Score 54; DB 18; Length 532;
XX	Best Local Similarity 100.0%; Pred. No. 1.9;
XX	Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	
XX	RESULT 14
XX	AAW22220
XX	AAW22220 standard; Protein: 537 AA.
XX	
XX	AAW22220;

```

XX 11-SEP-1997 (first entry)
XX
XX Protein encoded by pV8F construct.
XX
XX Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;
XX truncation; wild type; PCR; polymerase chain reaction; amplification;
XX proteolytic activity; fusion protein; beta-galactosidase; urea.
XX
XX Chimeric - Escherichia coli.
XX
XX Chimeric - Staphylococcus aureus.
XX
XX
XX Key Location/Qualifiers
XX Region 1..100
XX /note="E. coli beta-galactosidase"
XX Region 101..124
XX /note="R6 linker sequence"
XX Region 125..339
XX /note="truncated S. aureus V8 protease"
XX Region 342..365
XX /note="R6 linker sequence"
XX Region 366..537
XX /note="aminoglucoside 3'-phosphotransferase"
XX
XX EP745669-A2.
XX
XX 04-DEC-1996.
XX
XX 31-MAY-1996; 96EP-0303939.
XX
XX 02-JUN-1995; 95JP-0170086.
XX
XX (SUNR ) SUNTORX LTD.
XX
XX Ohshye K, Yabuta M;
XX
XX WPI; 1997-013693/02.
XX
XX Staphylococcus aureus V8 protease mutants - with increased
XX resistance to denaturation
XX
XX Claim 8; Page 19-20; 42pp; English.
XX
XX The invention relates to new mutant Staphylococcus aureus V8 proteases
XX which have enzyme activity even under environmental conditions which
XX promote protein denaturation. The mutants are based on 3 truncated V8
XX proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
XX acids from the C-terminal of the wild type protease. The mutants also
XX contain amino acid substitutions, especially D44E, N71S and/or R147K.
XX The protein sequence shown here represents a chimeric protein
XX comprising a truncated Staphylococcus aureus V8 protease lacking the
XX prepro and C-terminal 53 amino acids linked, via a synthetic R6 linker,
XX downstream of the E. coli beta-galactosidase. Also included downstream of
XX the V8 protease fragment is a second R6 linker and a fragment of the
XX aminoglucoside 3'-phosphotransferase protein. The chimeric sequence
XX was generated by restriction digestion and ligation from the V8RPT(-)
XX sequence (see AAW22218) by using a natural EcoRV site which removed a
XX further 8 amino acid from the C-terminus. This truncated V8 protease,
XX designated V8F, retains its level of activity in the presence of a
XX higher concentration of protein denaturant e.g. 5 M urea.
XX
XX Sequence 537 AA;
XX
XX Query Match 100.0%; Score 54; DB 18; Length 537;
XX Best Local Similarity 100.0%; Pred. No. 1.9;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 TCGNGSGSPVF 10
XX |||||||||
XX Db 289 TCGNGSGSPVF 298

```

```

AA015205
ID   AA015205 standard; Protein; 712 AA.
XX
AC   AA015205;
XX
DT   05-SEP-2002 (first entry)
XX
DE   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).
XX
KW   Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KM   DPP-7 inhibitor identification; periodontal disease; gingivitis;
XX   periodontitis.
XX
OS   Porphyromonas gingivalis.
XX
PN   WO200238742-A2.
XX
PD   16-MAY-2002.
XX
PF   08-NOV-2001; 2001WO-US46782.
XX
PR   08-NOV-2000; 2000US-246827P.
XX
PA   (UNGE-) UNIV GEORGIA RES FOUND INC.
XX
PI   Travis J, Potempa JS, Banbula A, Bugno M;
XX
DR   WPI: 2002-490075/52.
XX
NR   N-PSDB: AAL43635.
XX
PT   Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT   the dipeptidylpeptidase for protecting an animal from periodontal
PT   disease caused by Porphyromonas gingivalis
XX
PS   Claim 7; Fig 4; 65pp; English.
XX
CC   The invention comprises the amino acid and coding sequence of the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC   enzyme has amidolytic activity for cleavage of a peptide bond between the
CC   second and third amino acids from the N-terminal end of a target peptide.
CC   The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC   substituent on the alpha-carbon atom of the second amino acid from the
CC   N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC   useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC   reducing the growth of a bacterium and protecting an animal from a
CC   periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC   or periodontitis). The present amino acid sequence represents the
CC   Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the
CC   invention.
XX
SQ   Sequence 712 AA:

```

Query Match 100.0%; Score 54; DB 23; Length 712;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY   1 TCGNSGSPVF 10
      |||||||||
DB   644 TCGNSGSPVF 653

```

Search completed: December 20, 2002, 11:24:52
 Job time : 54.4737 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 08:33:52 ; Search time 26.8421 Seconds

(without alignments)
35.815 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TGGNSGSPVF 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	336	1	PRASAK
2	54	100.0	342	2	G89873
3	54	100.0	357	2	S21758
4	48	88.9	716	2	G82627
5	44	81.5	284	2	S25140
6	43	79.6	316	2	A45134
7	42	77.8	169	2	A12436
8	41	75.9	521	2	S62794
9	40	74.1	315	2	E71729
10	40	74.1	361	2	B56940
11	40	74.1	363	2	A56940
12	40	74.1	497	2	B97729
13	39	72.2	319	2	E87365
14	39	72.2	328	1	JN0453
15	39	72.2	364	2	AE3453
16	39	72.2	379	2	S23573
17	39	72.2	607	2	S01939
18	39	72.2	996	2	T50186
19	39	72.2	997	2	T19521
20	38	72.2	997	2	S63064
21	38	70.4	434	2	S23762
22	38	70.4	470	2	D75375
23	38	70.4	476	2	B82414
24	37	68.5	182	2	T49812
25	37	68.5	235	2	D89667
26	37	68.5	239	2	B89667
27	37	68.5	240	2	C89667
28	37	68.5	334	2	D82793
29	37	68.5	401	2	A48423

ALIGNMENTS

30	37	68.5	449	2	B71265	Probable Mg2+ tran
31	37	68.5	462	2	H97292	UDP-N-acetyluram
32	37	68.5	590	2	AB1411	autolysin, N-acety
33	37	68.5	817	2	T21336	hypothetical prote
34	37	68.5	830	2	T49270	receptor protein k
35	37	68.5	1254	2	T47141	hypothetical prote
36	37	68.5	1306	2	A70934	hypothetical glyci
37	37	68.5	1777	2	T34369	hypothetical prote
38	36	66.7	280	2	T28684	hypothetical prote
39	36	66.7	320	2	UC1311	cell protein precu
40	36	66.7	374	2	T09111	probable magnesium
41	36	66.7	381	2	JC7650	pectin lyase (EC 4
42	36	66.7	411	2	T15209	hypothetical prote
43	36	66.7	425	2	D64149	hypothetical prote
44	36	66.7	451	2	A82172	magnesium transpor
45	36	66.7	493	2	C97605	probable serine pr

RESULT 1

PRASAK
glutamyl endopeptidase (EC 3.4.21.19) precursor - Staphylococcus aureus
N:Alternate names: staphylococcal serine proteinase
C:Species: Staphylococcus aureus
C>Date: 04-Dec-1986 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1999
C:Accession: A26812; A00966
R:Carmona, C.; Gray, G.L.
Nucleic Acids Res. 15, 6757, 1987
A>Title: Nucleotide sequence of the serine protease gene of Staphylococcus aureus, st
A:Reference number: A26812; MUID:87316953; PMID:3306605
A:Accession: A26812
A:Molecule type: DNA
A:Residues: 1-336 <CAR>
A:Cross-references: EMBL:X00356; NID:g46686; PIDN:CAA68434.1; PID:g46687
A:Experimental source: strain V8
R:Drapeau, G.R.
Can. J. Biochem. 56, 534-544, 1978
A>Title: The primary structure of staphylococcal protease.
A:Reference number: A23824; MUID:78212487; PMID:96922
A:Accession: A00966
A:Molecule type: protein
A:Residues: 69-108;110-124;126-144, 'D', 146-192, 'T', 194-228, 'N', 230-258, 'Q', 260, 'D', 26
A:Experimental source: strain V8
C:Comment: This extracellular proteolytic enzyme preferentially cleaves the peptide b
C:Comment: This enzyme may be distantly related to the trypsin-type serine proteinase
dus.
C:Superfamily: staphylococcal serine proteinase
C:Keywords: hydrolase; serine proteinase
F:99336/Product: staphylococcal serine proteinase #status experimental <ACT>
F:119,161/Active site: His, Asp #status predicted
F:237/Active site: Ser #status experimental

Query Match 100.0%; Score 54; DB 1; Length 336;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGNSGSPVF 10
DB 233 TGGNSGSPVF 242

RESULT 2

G89873
hypothetical protein sspA [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: G89873
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Uli, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001

```

A:Authors: V. Rosa, A.J. de M., de Rosa Jr., V.E. de Sa, R.G. Santelli, R.V. Sawa  

A:Authors: da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva Jr., W.A., da Sil  

M.: Tsubako, M.H., Vailiada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L  

A:Reference number: A59328  

A:Contents: annotation  

C:Genetics:  

A:Gene: XF1887

Query Match      88.9%; Score 48; DB 2; Length 716;  

Best Local Similarity 100.0%; Pred. No. 2.8;  

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGNGSPV 9
      |||||
Db      647 TGGNGSPV 655

RESULT 5
S25140
serine proteinase homolog - Enterococcus faecalis
C:Species: Enterococcus faecalis
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-2000
C:Accession: S25140
R:Su, Y.A.; Clewell, D.B.
submitted to the EMBL Data Library, June 1992
A:Description: A gene (SPRE) downstream of gele of Enterococcus faecalis OG1-10 resen
A:Reference number: S25140
A:Accession: S25140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1284 <SVU>
A:Cross-references: EMBL:Z12296; NID:943337; PIDN:CAW8168.1; PID:943338
C:Superfamily: staphylococcal serine proteinase

Query Match      81.5%; Score 44; DB 2; Length 284;  

Best Local Similarity 70.0%; Pred. No. 5.2;  

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY      1 TGGNGSPVF 10
      ||| |||:
Db      224 TGGGSGSPVY 233

RESULT 6
A45134
endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A45134; S23078
R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S  

J. Biol. Chem. 267, 23782-23788, 1992
A:Title: Purification, characterization, cloning, and expression of a glutamic acid-e  

A:Reference number: A45134; MUID:93054737; PMID:1429718
A:Accession: A45134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <KRAK>
A:Cross-references: GB:D10060; NID:g216263; PIDN:BAA00949.1; PID:d1001415; PID:g216262  

A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBI:P:118785)  

R:Swendsen, I.; Bredam, K.  

Eur. J. Biochem. 204, 165-171, 1992
A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase  

A:Reference number: S23078; MUID:92155199; PMID:1346764
A:Accession: S23078
A:Status: preliminary
A:Molecule type: protein
A:Residues: 95-316 <SVE>
C:Keywords: hydrolase

Query Match      79.6%; Score 43; DB 2; Length 316;  

Best Local Similarity 88.9%; Pred. No. 8.5;  

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

A:Authors: V. Rosa, A.J. de M., de Rosa Jr., V.E. de Sa, R.G. Santelli, R.V. de Sa,
A:Authors: da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva Jr., W.A., da Silva
M.; Tsubako, M.H.; Vailiada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1887

Query Match      88.9%; Score 48; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 2.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 TGGNGSPV 9
      |||||
Db      647 TGGNGSPV 655

RESULT 5
S25140
serine proteinase homolog - Enterococcus faecalis
C:Species: Enterococcus faecalis
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-2000
C:Accession: S25140
R:Su, Y.A.; Clewell, D.B.
submitted to the EMBL Data Library, June 1992
A:Description: A gene (SPRE) downstream of gele of Enterococcus faecalis OG1-10 resen
A:Reference number: S25140
A:Accession: S25140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1284 <SVU>
A:Cross-references: EMBL:Z12296; NID:943337; PIDN:CAW8168.1; PID:943338
C:Superfamily: staphylococcal serine proteinase

Query Match      81.5%; Score 44; DB 2; Length 284;
Best Local Similarity 70.0%; Pred. No. 5.2;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY      1 TGGNGSPVF 10
      ||| |||
Db      224 TGGGSGSPVY 233

RESULT 6
A45134
endopeptidase (EC 3.4.-.-), glutamate-specific - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A45134; S23078
R:Kakudo, S.; Kikuchi, N.; Kitadokoro, K.; Fujiwara, T.; Nakamura, E.; Okamoto, H.; S
J. Biol. Chem. 267, 23782-23788, 1992
A:Title: Purification, characterization, cloning, and expression of a glutamic acid-e
A:Reference number: A45134; MUID:93054737; PMID:1429718
A:Accession: A45134
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-316 <KRAK>
A:Cross-references: GB:D10060; NID:g216263; PIDN:BAA00949.1; PID:d1001415; PID:g216262
A:Note: sequence extracted from NCBI backbone (NCBIN:118784, NCBI:P:118785)
R:Swendsen, I.; Bredam, K.
Eur. J. Biochem. 204, 165-171, 1992
A:Title: Isolation and amino acid sequence of a glutamic acid specific endopeptidase
A:Reference number: S23078; MUID:92155199; PMID:1346764
A:Accession: S23078
A:Status: preliminary
A:Molecule type: protein
A:Residues: 95-316 <SVE>
C:Keywords: hydrolase

Query Match      79.6%; Score 43; DB 2; Length 316;
Best Local Similarity 88.9%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 2 GGNSSGPVF 10
|||
Db 258 GGNSSGPVF 266

RESULT 7

AI2436
hypothetical protein alr5049 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AI2436

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AI2436

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-169 <KUR>

A:Cross-references: GB:BA000019; PIDN:BAH76748.1; PID:g17134187; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr5049

Query Match 77.8%; Score 42; DB 2; Length 169;
Best Local Similarity 88.9%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNSSGPV 9
|||||
Db 109 TGNSSASPV 117

RESULT 8

S62794
probable lipoprotein D02_orf521 - Mycoplasma pneumoniae (strain ATCC 29342)

N:Alternate names: MG395 homolog D02_orf521

C:Species: Mycoplasma pneumoniae

A:Variety: ATCC 29342

C:date: 27-Apr-1996 #sequence_revision 01-Nov-1996 #text_change 07-Dec-1999

C:Accession: S62794; S62798; S73576

R:Himmelfreich, R.; Hilbert, H.

submitted to the EMBL Data Library, December 1995

A:Reference number: S62794

A:Accession: S62794

A:Molecule type: DNA

A:Residues: 1-521 <HIM>

A:Cross-references: EMBL:U43738; NID:g1209757; PIDN:AAC43664.1; PID:g1209771

R:Hilbert, H.; Himmelfreich, R.; Plagens, H.; Herrmann, R.

Nucleic Acids Res. 24, 628-639, 1996

A:title: Sequence analysis of 56 kb from the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S62797; MUID:96177562; PMID:8604030

A:Accession: S62798

A:Molecule type: DNA

A:Residues: 1-50 <HIL>

A:Cross-references: EMBL:U43738

R:Himmelfreich, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li, B.C.; Herrmann, R.

Nucleic Acids Res. 24, 4420-4449, 1996

A:title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae

A:Reference number: S73327; MUID:97105885; PMID:8948633

A:Accession: S73576

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-521 <HIL>

A:Cross-references: EMBL:AE000025; GB:U00089; NID:g1673918; PIDN:AAB95899.1; PID:g167391

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996

C:Genetics:

A:Genetic code: SGC3

C:Superfamily: hypothetical protein MG068

Query Match 75.9%; Score 41; DB 2; Length 521;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGNSSGPVF 10
|||
Db 423 GGNSSGPVF 431

RESULT 9

E71729
proteinnase DO (htra) RP186 - Rickettsia prowazekii

C:Species: Rickettsia prowazekii

C:date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000

C:Accession: E71729

R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark

Nature 396, 133-140, 1998

A:title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.

A:Reference number: A71630; MUID:99039499; PMID:9823893

A:Accession: E71729

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-315 <AND>

A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14652.1; PID:e134

A:Experimental source: strain Madrid E

C:Genetics:

A:Gene: htra; RP186

Query Match 74.1%; Score 40; DB 2; Length 315;
Best Local Similarity 88.9%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNSSGPV 9
|||||
Db 188 TGNSSGPV 196

RESULT 10

B56940
Integral membrane protein TGN38A - mouse

C:Species: Mus musculus (house mouse)

C:date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995

C:Accession: B56940

R:Kasai, K.; Takahashi, S.; Murakami, K.; Nakayama, K.

J. Biol. Chem. 270, 14471-14476, 1995

A:title: Strain-specific presence of two TGN38 isoforms and absence of TGN41 in mouse

A:Reference number: A56940; MUID:95301533; PMID:7540170

A:Accession: B56940

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-361 <KAS>

A:Cross-references: GB:D50032

Query Match 74.1%; Score 40; DB 2; Length 361;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNSSGP 8
|||||
Db 138 TGNSSGP 145

RESULT 11

A56940

Integral membrane protein TGN38B - mouse

C:Species: Mus musculus (house mouse)

C:date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995

C:Accession: A56940

R:Kasai, K.; Takahashi, S.; Murakami, K.; Nakayama, K.

J. Biol. Chem. 270, 14471-14476, 1995

A:title: Strain-specific presence of two TGN38 isoforms and absence of TGN41 in mouse

A:Reference number: A56940; MUID:95301533; PMID:7540170

A:Accession: A56940

A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-363 <KAS>
A:Cross-references: GB:ID50031

Query Match 74.1%; Score 40; DB 2; Length 363;
Best Local Similarity 87.5%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 8
||| ||| |||
DB 140 TCGNSGSPV 147

RESULT 12

B97729
heat shock proteinase (EC 3.4.21.-) [imported] - Rickettsia conorii (strain Malish 7)

C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001

C:Accession: B97729

R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fourmter, P.E.; Barbe, V.; Samson, D.; R
Science 293, 2093-2098, 2001

A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893

A:Accession: B97729

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-497 <KUR>

A:Cross-references: GB:AE006914; PIDN:AAL02772.1; PID:g15619287; GSPDB:GN00173

C:Genetics:

A:Gene: RC0234

C:Keywords: hydrolase; serine proteinase

Query Match 74.1%; Score 40; DB 2; Length 497;
Best Local Similarity 88.9%; Pred. No. 43;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
||| ||| |||
DB 188 TCGNSGSPV 196

RESULT 13

E87365

hypothetical protein CC0937 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: E87365

R:Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DebRoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
N.; J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: E87365

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-319 <STO>

A:Cross-references: GB:AE005673; NID:g13422209; PIDN:AAK2921.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0937

Query Match 72.2%; Score 39; DB 2; Length 319;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 GNSGSPVF 10
||| ||| |||
DB 222 GNSGSPVF 229

RESULT 14

JN0453

dihydroorotate oxidase (EC 1.3.3.1) - g11 mushroom (Agrocycbe aegerita)

N:Alternate names: dihydroorotate dehydrogenase

C:Species: Agrocycbe aegerita

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: JN0453

R:Noel, T.; Labarete, J.

Gene 122, 233-234, 1992

A:Title: Sequence of the URA1 gene encoding dihydroorotate dehydrogenase from the bas

A:Reference number: JN0453; MUID:93083991; PMID:1452035

A:Accession: JN0453

A:Molecule type: DNA

A:Residues: 1-328 <NOE>

A:Cross-references: GB:M90295; NID:g166337; PIDN:AAA32636.1; PID:g166338

C:Genetics:

A:Gene: URA1

C:Superfamily: dihydroorotate oxidase

C:Keywords: Flavoprotein; oxidoreductase; pyrimidine nucleotide biosynthesis

Query Match 72.2%; Score 39; DB 1; Length 328;
Best Local Similarity 70.0%; Pred. No. 41;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10
||| ||| |||
DB 261 TCGNSGSPVF 270

RESULT 15

AE3453

dihydroorotate oxidase (EC 1.3.3.1) [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 19-Apr-2002

C:Accession: AE3453

R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov

.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit

A:Reference number: AD3252; PMID:11756688

A:Accession: AE3453

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-364 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAL52792.1; PID:g17983628; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BME11611

A:Map position: 1

C:Superfamily: dihydroorotate oxidase

C:Keywords: oxidoreductase

Query Match 72.2%; Score 39; DB 2; Length 364;
Best Local Similarity 70.0%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10
||| ||| |||
DB 261 TCGNSGSPVF 270

Search completed: December 20, 2002, 12:11:30
Job time : 34.8421 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 11:25:03 ; Search time 12.6316 Seconds
(without alignments)
13.271 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TCGNSGSPVF 10

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications-AA:*

1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCT07_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	10	US-10-008-355-25	Sequence 25, Appl
2	54	100.0	52	US-10-008-355-3	Sequence 3, Appl
3	54	100.0	52	US-10-008-355-4	Sequence 4, Appl
4	54	100.0	699	US-10-008-355-8	Sequence 8, Appl
5	54	100.0	712	US-10-008-355-2	Sequence 2, Appl
6	54	100.0	9	US-10-008-355-6	Sequence 6, Appl
7	48	88.9	9	US-10-008-355-26	Sequence 26, Appl
8	48	88.9	716	US-10-008-355-7	Sequence 7, Appl
9	48	88.9	720	US-10-008-355-9	Sequence 9, Appl
10	44	81.5	734	US-10-008-355-5	Sequence 5, Appl
11	37	68.5	74	US-09-864-761-48349	Sequence 48349, A
12	37	68.5	240	US-09-815-242-5351	Sequence 5351, A
13	37	68.5	254	US-09-815-242-12277	Sequence 12277, A
14	36	66.7	447	US-09-388-089B-2	Sequence 2, Appl
15	36	66.7	475	US-09-388-089B-12	Sequence 12, Appl
16	36	66.7	498	US-09-388-089B-11	Sequence 11, Appl
17	35	64.8	499	US-09-864-761-35385	Sequence 35385, A
18	35	64.8	503	US-09-752-385-8	Sequence 8, Appl
19	35	64.8	550	US-09-905-657-2	Sequence 2, Appl

20	35	64.8	550	10	US-09-883-797-8	Sequence 8, Appl
21	34	63.0	77	10	US-09-764-869-976	Sequence 976, App
22	33	61.1	40	10	US-09-864-761-34778	Sequence 34778, A
23	33	61.1	74	10	US-09-864-761-42729	Sequence 42729, A
24	33	61.1	223	12	US-10-003-152-4	Sequence 4, Appl
25	33	61.1	440	9	US-10-066-500-106	Sequence 106, App
26	33	61.1	440	9	US-10-063-547-52	Sequence 52, Appl
27	33	61.1	440	12	US-10-006-867-52	Sequence 52, Appl
28	33	61.1	440	12	US-10-052-586-202	Sequence 202, App
29	33	61.1	656	10	US-09-728-910-4	Sequence 4, Appl
30	33	61.1	660	10	US-09-728-910-2	Sequence 2, Appl
31	33	61.1	1070	10	US-09-735-367B-6	Sequence 6, Appl
32	33	61.1	1981	9	US-09-928-457-38	Sequence 38, Appl
33	33	61.1	2005	10	US-09-735-367B-3	Sequence 3, Appl
34	33	61.1	2063	10	US-09-735-367B-2	Sequence 2, Appl
35	32	59.3	43	10	US-09-864-761-40613	Sequence 40613, A
36	32	59.3	56	10	US-09-864-761-35910	Sequence 35910, A
37	32	59.3	94	10	US-09-864-761-34548	Sequence 34548, A
38	32	59.3	97	10	US-09-734-017A-40	Sequence 40, Appl
39	32	59.3	180	9	US-10-023-182-8	Sequence 8, Appl
40	32	59.3	180	10	US-09-751-798-8	Sequence 8, Appl
41	32	59.3	230	10	US-09-815-242-13533	Sequence 13533, A
42	32	59.3	351	10	US-09-853-625B-16	Sequence 16, Appl
43	32	59.3	371	10	US-09-739-861A-1	Sequence 1, Appl
44	32	59.3	371	10	US-09-795-583-1	Sequence 1, Appl
45	32	59.3	386	10	US-09-739-861A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1

US-10-008-355-25

Sequence 25, Application US/10008355

Patent No. US20020164759A1

GENERAL INFORMATION:

APPLICANT: Travis, James

APPLICANT: Potempa, Jan S

APPLICANT: Banbula, Agnieszka

TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

FILE REFERENCE: 235.00440101

CURRENT FILING DATE: 2001-11-08

PRIOR APPLICATION NUMBER: US 60/246,827

PRIOR FILING DATE: 2000-11-08

NUMBER OF SEQ ID NOS: 26

SOFTWARE: PatentIn version 3.0

SEQ ID NO 25

LENGTH: 10

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Consensus sequence

US-10-008-355-25

Query Match

Best Local Similarity 100.0%; Score 54; DB 9; Length 10;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10

Db 1 TCGNSGSPVF 10

RESULT 2

US-10-008-355-3

Sequence 3, Application US/10008355

Patent No. US20020164759A1

GENERAL INFORMATION:

APPLICANT: Travis, James

APPLICANT: Potempa, Jan S

APPLICANT: Banbula, Agnieszka

TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

```
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin version 3.0
SEQ ID NO 3
LENGTH: 52
TYPE: PRT
ORGANISM: Porphyromonas gingivalis
US-10-008-355-3
```

```
Query Match          100.0%; Score 54; DB 9; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TGCNCGSPVF 10
    |||
Db 1 TGCNCGSPVF 10
```

APP

```
RESULT 3
US-10-008-355-4
Sequence 4, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin version 3.0
SEQ ID NO 4
LENGTH: 52
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-10-008-355-4
```

```
Query Match          100.0%; Score 54; DB 9; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TGCNCGSPVF 10
    |||
Db 1 TGCNCGSPVF 10
```

APP

```
RESULT 4
US-10-008-355-8
Sequence 8, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin version 3.0
SEQ ID NO 8
LENGTH: 699
TYPE: PRT
ORGANISM: Porphyromonas gingivalis
```

```
US-10-008-355-8
```

```
Query Match          100.0%; Score 54; DB 9; Length 699;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TGCNCGSPVF 10
    |||
Db 644 TGCNCGSPVF 653
```

APP

```
RESULT 5
US-10-008-355-2
Sequence 2, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin version 3.0
SEQ ID NO 2
LENGTH: 712
TYPE: PRT
ORGANISM: Porphyromonas gingivalis
US-10-008-355-2
```

```
Query Match          100.0%; Score 54; DB 9; Length 712;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TGCNCGSPVF 10
    |||
Db 644 TGCNCGSPVF 653
```

APP

```
RESULT 6
US-10-008-355-6
Sequence 6, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentin version 3.0
SEQ ID NO 6
LENGTH: 732
TYPE: PRT
ORGANISM: Shewanella putrefaciens
US-10-008-355-6
```

```
Query Match          100.0%; Score 54; DB 9; Length 732;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 TGCNCGSPVF 10
    |||
Db 662 TGCNCGSPVF 671
```


RESULT 7
US-10-008-355-26
; Sequence 26, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus sequence
US-10-008-355-26

Query Match 88.9%; Score 48; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 8.9e+04;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 9
|||||
Db 1 TGNSSGSPV 9

RESULT 8
US-10-008-355-7
; Sequence 7, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-008-355-7

Query Match 88.9%; Score 48; DB 9; Length 716;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 9
|||||
Db 647 TGNSSGSPV 655

RESULT 9
US-10-008-355-9
; Sequence 9, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-9

Query Match 88.9%; Score 48; DB 9; Length 720;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 9
|||||
Db 651 TGNSSGSPV 659

RESULT 10
US-10-008-355-5
; Sequence 5, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-5

Query Match 81.5%; Score 44; DB 9; Length 734;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGSP 8
|||||
Db 665 TGNSSGSP 672

RESULT 11
US-09-664-761-48349
; Sequence 48349, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03

```
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48349
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121716.16
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EST HUMAN HIT: BE002805.1, EVALU8 8.00e-39
; OTHER INFORMATION: SWISSPROT HIT: Q92KD2, EVALU8 3.60e+00
US-09-864-761-48349

Query Match          68.5%; Score 37; DB 10; Length 74;
Best Local Similarity 75.0%; Pred. No. 9.3;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 GNGSGSPV 9
DB      52 GNGSGSPI 59

RESULT 12
US-09-815-242-5351
; Sequence 5351, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

```
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5351
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5351

Query Match          68.5%; Score 37; DB 10; Length 240;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 GNGSGSPV 9
DB      191 GNGSGSPV 197

RESULT 13
US-09-815-242-12277
; Sequence 12277, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12277
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12277

Query Match          68.5%; Score 37; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 GNGSGSPV 9
```

Db 205 GNSGSPV 211

RESULT 14
US-09-388-089B-2
; Sequence 2, Application US/09388089B
; Patent No. US20020018782A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, W.
; APPLICANT: Harris, A.
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID
; FILE REFERENCE: 7969-083
; CURRENT APPLICATION NUMBER: US/09/388,089B
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Neisseria spp.
US-09-388-089B-2

Query Match 66.7%; Score 36; DB 10; Length 447;
Best Local Similarity 75.0%; Pred. No. 82;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GNSGSPV 10
|||||:|
Db 192 GNSGSPV 199

RESULT 15
US-09-388-089B-12
; Sequence 12, Application US/09388089B
; Patent No. US20020018782A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, W.
; APPLICANT: Harris, A.
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID
; FILE REFERENCE: 7969-083
; CURRENT APPLICATION NUMBER: US/09/388,089B
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-388-089B-12

Query Match 66.7%; Score 36; DB 10; Length 475;
Best Local Similarity 75.0%; Pred. No. 88;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 GNSGSPV 10
|||||:|
Db 220 GNSGSPV 227

Search completed: December 20, 2002, 12:17:26
Job time : 13.6316 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 05:38:27 ; Search time 15.2632 Seconds
(without alignments)
27.174 Million cell updates/sec

Title: US-10-008-355-25
Perfect score: 54
Sequence: 1 TCGNSGSPVF 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54	100.0	336	1 STSP_STAAU	P0188 staphylococ
2	43	79.6	316	1 GSEP_BACLI	P80057 bacillus li
3	42	77.8	418	1 ER24_HUMAN	O76062 homo sapien
4	41	75.9	521	1 YF92_MYCPN	O50335 mycoplasma
5	39	72.2	328	1 PYRD_AGRAB	P28294 agrocyste ae
6	39	72.2	378	1 PLYB_ASPPG	O00205 aspergillus
7	39	72.2	607	1 V66K_BWYVF	P09506 beet wester
8	39	72.2	997	1 YNN3_YEAST	P53920 saccharomyc
9	39	72.2	1035	1 RRPQ_BWYVF	P09507 beet wester
10	38	70.4	434	1 PYRD_ARATH	P32746 arabidopsis
11	37	68.5	401	1 HME1_MOUSE	P09065 mus musculu
12	36	66.7	320	1 CE1L_AGABI	O00023 agaricus bi
13	36	66.7	380	1 PIRA_COLGL	O00374 colletoctic
14	36	66.7	425	1 Y350_HAEIN	P24326 haemophilus
15	36	66.7	508	1 DEGP_RICCN	O92361 rickettsia
16	36	66.7	513	1 DEGP_RICPR	O05942 rickettsia
17	36	66.7	516	1 Y067_MYCGE	P47313 mycoplasma
18	36	66.7	628	1 ABFA_ASPPG	P42254 aspergillus
19	36	66.7	921	1 ITH4_PIG	P79263 sus scrofa
20	35	64.8	108	1 Y108_NPVOP	O10347 oryza pisen
21	35	64.8	417	1 IRX5_HUMAN	P74411 homo sapien
22	35	64.8	503	1 DEGP_BARHE	P54925 bartonella
23	35	64.8	504	1 DEGP_RHME	O58924 bartonella
24	35	64.8	513	1 DEGP_BRUAB	O44597 bruceella ab
25	35	64.8	524	1 Y395_MYCGE	P47635 mycoplasma
26	35	64.8	531	1 YF88_MYCPN	O50339 mycoplasma
27	35	64.8	583	1 YAB3_MYCPN	P75610 mycoplasma
28	35	64.8	584	1 YMB3_YEAST	O04228 saccharomyc
29	35	64.8	1394	1 HAP1_HAEIN	P45387 haemophilus
30	35	64.8	1409	1 HAP1_HAEIN	P45387 haemophilus
31	35	64.8	1532	1 IGA1_NITGO	P09790 neisseria g
32	35	64.8	1541	1 IGA1_HAEIN	P43782 haemophilus
33	35	64.8	1545	1 IGA3_HAEIN	P45385 haemophilus

34	35	64.8	1694	1 IGA0_HAEIN	P44969 haemophilus
35	35	64.8	1702	1 IGA2_HAEIN	P45384 haemophilus
36	35	64.8	1714	1 SYEP_DROME	P26668 drosophila
37	35	64.8	1849	1 IGA4_HAEIN	P45386 haemophilus
38	34	63.0	154	1 KRSC_CHICK	P04459 gallus gall
39	34	63.0	157	1 YF89_MYCPN	O50338 mycoplasma
40	34	63.0	303	1 Y007_MYCLE	O32870 mycobacteri
41	34	63.0	325	1 YF65_STRPN	O97pn8 streptococc
42	34	63.0	353	1 YF91_MYCPN	O50336 mycoplasma
43	34	63.0	424	1 CBPT_THERVD	P29068 thermactin
44	34	63.0	485	1 CAT1_NICPL	P49315 nicotiana p
45	34	63.0	492	1 CAT2_LYCES	O9xnh3 lycopersico

ALIGNMENTS

```

RESULT 1
ID STSP_STAAU STANDARD: PRT: 336 AA.
AC P0188:
  20-WAR-1987 (rel. 04, Created)
  01-APR-1988 (rel. 07, Last sequence update)
  15-JUN-2002 (rel. 41, Last annotation update)
  DE Glutamyl endopeptidase precursor (EC 3.4.21.19) (Staphylococcal serine
  DE proteinase) (V8 proteinase) (Endoproteinase Glu-C).
  GN SSPA.
  OS Staphylococcus aureus.
  OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
  NCBI_TaxID=1280;
  [1]
  RP SEQUENCE FROM N.A.
  RC STRAIN-V8:
  RX MEDLINE=87316953; PubMed=3306605;
  RA Carmona C., Gray G.L.;
  RT "Nucleotide sequence of the serine protease gene of Staphylococcus
  RT aureus, strain V8.";
  RL Nucleic Acids Res. 15:6757-6757(1987).
  RN [2]
  RP SEQUENCE FROM N.A.
  RX MEDLINE=20569178; PubMed=11119502;
  RA Rice K., Peralta R., Bast D., de Azavedo J., McGavin M.J.;
  RT "Description of staphylococcus serine protease (ssp) operon in
  RT staphylococcus aureus and nonpolar inactivation of sspA-encoded serine
  RT protease.";
  RL Infect. Immun. 69:159-169(2001).
  RN [3]
  RP SEQUENCE OF 69-280.
  RC STRAIN-V8:
  RX MEDLINE=78212487; PubMed=96922;
  RA Drapeau G.R.;
  RT "The primary structure of staphylococcal protease.";
  RL Can. J. Biochem. 56:534-544(1978).
  CC -I- FUNCTION: PREFERENTIALLY CLEAVES PEPTIDE BONDS ON THE CARBOXYL-
  CC TERMINAL SIDE OF ASPARTATE AND GLUTAMATE.
  CC -I- CATALYTIC ACTIVITY: Preferential cleavage: Asp-I-Xaa, Glu-I-Xaa.
  CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.
  CC -I- DATABASE: NAME=Washington enzyme manual;
  CC WWW="http://www.washington-biochem.com/manual/P/STAP.html".
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@sib-sib.ch).
  CC -----
  DR EMBL: Y00365; CA68434.1; -
  DR EMBL: AF309515; AAG45843.1; -
  DR PIR: A26812; PRSASK.
  DR MEROPS: S01.269; -
  DR InterPro: IPR000126; Ser_proteas_V8.

```

```
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin. 1.
DR PRINTS: PR00839; V8PROTEASE.
DR PROSITE: PS00672; V8_HIS. 1.
DR PROSITE: PS00673; V8_SER. 1.
KM Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 29
FT PROPEP 30 68
FT CHAIN 69 336
FT ACT_SITE 119 119
FT ACT_SITE 161 161
FT ACT_SITE 237 237
FT ACT_SITE 109 109
FT CONFLICT 125 125
FT CONFLICT 145 145
FT CONFLICT 193 193
FT CONFLICT 229 229
FT CONFLICT 259 261
FT CONFLICT 268 270
SQ SEQUENCE 336 AA; 36326 MW; 8B138D0C7996AA3E CRC64;

Query Match
Best Local Similarity 100.0%; Score 54; DB 1; Length 336;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGSGSPVF 10
DB 233 TCGSGSPVF 242

RESULT 2
GSEP_BACLI
ID GSEP_BACLI STANDARD; PRT; 316 AA.
AC P80057;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamyl endopeptidase precursor (EC 3.4.21.19) (Glutamate specific
DE endopeptidase) (GSE).
CN BLASE.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 14580;
RX MEDLINE=93054737; PubMed=1429718;
RA Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,
RA Okamoto H., Shin M., Tamaki M., Teraoka H., Tsuzuki H., Yoshida N.;
RT "Purification, characterization, cloning, and expression of a
RT glutamic acid-specific protease from Bacillus licheniformis ATCC
RT 14580.";
RL J. Biol. Chem. 267:23782-23788(1992).
RN [2]
RP SEQUENCE OF 95-316.
RX MEDLINE=9215519; PubMed=1346764;
RA Svendsen I., Bredam K.;
RT "Isolation and amino acid sequence of a glutamic acid specific
RT endopeptidase from Bacillus licheniformis.";
RL Eur. J. Biochem. 204:165-171(1992).
CC -!- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE
CC CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG
CC PREFERENCE FOR GLU.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Asp-I'-Xaa, Glu-I'-Xaa.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
```

```
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D10060; BAA00949.1; -.
CC DR PIR: S23078; S23078.
CC DR PIR: A45134; A45134.
CC DR MEROPS: S01.271; -.
DR InterPro: IPR001254; Ser_protease_V8.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin. 1.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00020; TRYP_SPC. 1.
DR PROSITE: PS00672; V8_HIS. 1.
DR PROSITE: PS00673; V8_SER. 1.
KM Hydrolyase; Serine protease; Signal.
FT SIGNAL 1 30
FT PROPEP 31 94
FT CHAIN 95 316
FT ACT_SITE 141 141
FT ACT_SITE 261 261
FT DISULFID 126 142
FT DISULFID 275 279
SQ SEQUENCE 316 AA; 33611 MW; 96D7552CB7089B09 CRC64;

Query Match
Best Local Similarity 79.6%; Score 43; DB 1; Length 316;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GGNSSGPVF 10
DB 258 GGNSSGPVF 266

RESULT 3
ER24_HUMAN
ID ER24_HUMAN STANDARD; PRT; 418 AA.
AC 076062; O95982; Q96G21; Q96G64;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Delta(14)-sterol reductase (EC 1.3.1.70) (C-14 sterol reductase)
DE (Sterol C14-reductase) (Delta14-SR) (Transmembrane 7 superfamily
DE member 2) (Another new gene 1) (Putative sterol reductase SR-1).
CN TM7SF2 OR ANG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98277456; PubMed=9615229;
RA Lemmens I.H., Kas K., Merregaert J., Van De Ven W.J.M.;
RT "Identification and molecular characterization of TM7SF2 in the FAUNA
RT gene cluster on human Chromosome 11q13.";
RL Genomics 49:437-442(1998).
RN [2]
RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
RX MEDLINE=99097347; PubMed=9878250;
RA Holmer L., Pezhan A., Worman H.U.;
RT "The human lamin B receptor/sterol reductase multigene family.";
RL Genomics 54:469-476(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Eye;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP FUNCTION.
RX PubMed=11784322;
RA Roberti R., Benatti A.M., Galli G., Caruso D., Matras B., Aisa C.,
RA Becarci T., Della Fazio M.A., Servillo G.;
RT "Cloning and expression of sterol Delta14-reductase from bovine
RT liver.";
RL Eur. J. Biochem. 269:283-290(2002).
```

CC -i- FUNCTION: Involved in the conversion of lanosterol to cholesterol.
CC -i- CATALYTIC ACTIVITY: 4,4-dimethyl-5-alpha-cholesta-8,24-dien-3-beta-ol + NADP(+) = 4,4-dimethyl-5-alpha-cholesta-8,14,24-trien-3-beta-ol + NADPH.
CC -i- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.
CC -i- TISSUE SPECIFICITY: Expressed in adult heart, brain, pancreas, lung, liver, skeletal muscle, kidney, ovary, prostate, and testis, but not detected in placenta, spleen, thymus, small intestine, colon (mucosal lining), or peripheral blood leukocytes.
CC -i- SIMILARITY: BELONGS TO THE ERG4/ERG24 FAMILY.
CC -i- CAUTION: Ref.1 sequence differs from that shown due to a number of sequencing problems as reported in Ref.2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF048704; AAC21457.1; ALT_FRAME.
CC EMBL: AF023676; AAC21450.1; ALT_FRAME.
CC EMBL: AF096303; AAD09769.1; -.
CC EMBL: AF096304; AAD09765.1; -.
CC EMBL: BC009052; AAH09052.1; -.
CC EMBL: BC012857; AAH12857.1; -.
CC Genew: HGNC:11863; TM/SF2.
CC MIM: 603414; -.
CC InterPro: IPR001171; ERG4_ERG24.
CC Pfam: PF01222; ERG4_ERG24; 1.
CC PROSITE: PS01017; STEROL_REDUCT_1; 1.
CC PROSITE: PS01018; STEROL_REDUCT_2; 1.
CC KM Sterol biosynthesis; Oxidoreductase; NADP; Transmembrane; Endoplasmic reticulum; Polymorphism.
CC FT TRANSMEM 13 35 POTENTIAL.
CC FT TRANSMEM 62 81 POTENTIAL.
CC FT TRANSMEM 102 124 POTENTIAL.
CC FT TRANSMEM 129 148 POTENTIAL.
CC FT TRANSMEM 255 277 POTENTIAL.
CC FT TRANSMEM 287 304 POTENTIAL.
CC FT TRANSMEM 355 377 POTENTIAL.
CC FT VARIANT 299 377 I -> T.
CC FT VARIANT 299 377 I -> T.
CC FT CONFLICT 179 179 /FTID=VAR.012716.
CC FT SEQUENCE 418 AA; 46417 MW; 357C8ABE2BEDA918 CRC64;
CC
CC Query Match 77.8%; Score 42; DB 1; Length 418;
CC Best Local Similarity 66.7%; Pred. No. 4.3;
CC Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

RA Hilbert H., Himmelreich R., Plagens H., Herrmann R.;
RT "Sequence analysis of 56 kb from the genome of the bacterium Mycoplasma pneumoniae comprising the dna region, the *atp* operon and a cluster of ribosomal protein genes";
RL Nucleic Acids Res. 24:628-639(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C., Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -i- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor (Potential).
CC -i- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: UA3738; AAC43664.1; -.
CC EMBL: AE000025; AAB93898.1; -.
CC InterPro: IPR002414; DUF30/31.
CC Pfam: PF01727; DUF30; 1.
CC Pfam: PF01732; DUF31; 1.
CC PRINTS: PR00840; Y06768FAMILY.
CC KM Hypothetical protein; Lipoprotein; Membrane; Signal; Complete proteome.
CC FT SIGNAL 1 22 POTENTIAL.
CC FT CHAIN 23 521 HYPOTHETICAL LIPOPROTEIN MPN592.
CC FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
CC SEQUENCE 521 AA; 59500 MW; 0E706CDE8CEEDA CRC64;
CC
CC Query Match 75.9%; Score 41; DB 1; Length 521;
CC Best Local Similarity 77.8%; Pred. No. 8.1;
CC Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 5
PYRD_AGRAE
ID PYRD_AGRAE STANDARD; PRT; 328 AA.
AC P28294;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dihydroorotate dehydrogenase (EC 1.3.3.1) (Dihydroorotate oxidase) (DHODase) (DHODase) (DHOD).
GN URA1.
OS Agrocycbe aegerita.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes; Agaricales; Boletifacaeae; Agrocycbe.
OX NCBI_TaxID=5400;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93083991; PubMed=1452035;
RA Noel T., Labatere J.;
RT "Sequence of the URA1 gene encoding dihydroorotate dehydrogenase from the basidiomycete fungus *Agrocycbe aegerita*.";
RL Gene 122:233-234(1992).
CC -i- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate + H(2)O(2).
CC -i- COFACTOR: FAD.
CC -i- PATHWAY: Pyrimidine biosynthesis; fourth step.

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M80295; AAA32636.1; -
DR PIR: JN0453; JN0453.
DR InterPro: IPR001295; DHO_dh.
DR Pfam: PF01180; DHodense; 1.
DR TIGRFAMs: TIGR01036; pyrd_sub2; 1.
DR PROSITE: PS00911; DHODEHASE_1; 1.
DR PROSITE: PS00912; DHODEHASE_2; 1.
DR Pyrimidine biosynthesis: Oxidoreductase; Flavoprotein; FAD.
KM NP_BIND 286 296 FAD (NAD PART) (POTENTIAL).
FT SEQUENCE 328 AA; 35085 MW; 71FA3D2AD57EF75 CRC64;
SQ
Query Match 72.2%; Score 39; DB 1; Length 328;
Best Local Similarity 70.0%; Pred. No. 11;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 TGGNSGSPVF 10
DB 261 TGGLSGAPLF 270

```

```

FT ACT_SITE 255 255 POTENTIAL.
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 376 AA; 39703 MW; 4FF321AF280B72FF CRC64;
Query Match 72.2%; Score 39; DB 1; Length 378;
Best Local Similarity 70.0%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGGNSGSPVF 10
DB 35 TGGGSASPYF 44

```

```

RESULT 6
ID PLYB_ASPNG STANDARD: PRT; 378 AA.
AC 000205;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Pectin lyase B precursor (EC 4.2.2.10) (PLB).
GN PELB.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 120.49 / N400;
RX MEDLINE=92357005; PubMed=1495474;
RA Kusters-Van Someren M., Filippi M., de Graaff L., van den Broeck H.,
RA Kester H., Himmen A., Visser J.;
RT "Characterization of the Aspergillus niger pelb gene: structure and
RT regulation of expression."
RL Mol. Gen. Genet. 234:113-120(1992).
CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectin to give
CC oligosaccharides with terminal 4-deoxy-6-methyl-alpha-D-galact-4-
CC enunosyl groups.
CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X65552; CAA6521.1; -
DR EMBL: A12248; CAA01023.1; -
DR HSSP: Q01172; 1IDJ.
DR InterPro: IPR002022; Amb.allergen.
DR Pfam: PF00544; pec_lyase; 1.
KM Lyase; Signal; Glycoprotein; Multigene family.
FT SIGNAL 1 20 OR 21 (POTENTIAL).
FT CHAIN 21 378 PECTIN LYASE B.

```

```

RESULT 7
ID V66K_BMYVF STANDARD: PRT; 607 AA.
AC P09506;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-OCT-1989 (Rel. 12, Last annotation update)
DE 66.2 kDa protein (ORF 2).
OS Beet western yellows virus (isolate FL-1) (BMYV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Polerovirus.
OX NCBI_TaxID=12043;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89057523; PubMed=3194229;
RA Veidt I., Lot H., Leiser M., Scheidecker D., Guillely H., Richards K.,
RA Jonard G.;
RT "Nucleotide sequence of beet western yellows virus RNA."
RL Nucleic Acids Res. 16:9917-9932(1988).
CC -1- SIMILARITY: TO POTATO LEAFROLL VIRUS ORF2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X13063; CAA31463.1; -
DR PIR: S01939; S01939.
DR MEROPS: S52.001; -.
DR InterPro: IPR000382; Luteo_ORF2.
DR Pfam: PF02122; Luteo_ORF2; 1.
DR PRINTS: PR00913; LVIRUSORF2.
SQ SEQUENCE 607 AA; 66210 MW; 2A41F82911DEC84F CRC64;
Query Match 72.2%; Score 39; DB 1; Length 607;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 GGNSSGSPVF 10
DB 354 GGHSGSPVF 362

```

```

RESULT 8
ID YNM3_YEAST STANDARD: PRT; 997 AA.
AC P53920;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical 110.9 kDa protein in SPC98-TOM70 intergenic region.
GN YNL123W OR N1897.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;

```



```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97245296; PubMed=9090055;
RA de Antoni A., D'Angelo M., Dal Pero F., Sartorello F., Pandolfo D.,
RA Pallavicini A., Lanfranchi G., Valle G.;
RT "The DNA sequence of cosmid 14-13b from chromosome XIV of
RT Saccharomyces cerevisiae reveals an unusually high number of
RT overlapping open reading frames."
RL Yeast 13:261-266(1997).
CC -1- SIMILARITY: TO S.POMBE SPBC1685.05.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: 269382; CAA93384.1; -
CC EMBL: 271399; CAA96004.1; -
CC MEROPS: S01 UPC; -
CC SGD: S0005067; YNLI23W.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001254; Ser_protease_Try.
CC Pfam: PF00595; PDZ; 1.
CC SMART: SM00228; PDZ; 2.
CC Hypothetical protein.
CC KW SEQUENCE 997 AA; 110881 MW; A26005C1DDDB932C CRC64;
SQ
Query Match 72.2%; Score 39; DB 1; Length 997;
Best local Similarity 77.8%; Pred. No. 36;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGGSGSPV 9
DB 232 GGGSGSPV 240
-----
RESULT 9
RPO_BWVF STANDARD; PRT; 1035 AA.
AC P09507;
DT 01-MAR-1989 (Rel. 10, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative RNA-directed RNA polymerase (EC 2.7.7.48) (ORF 3).
OS Beet western yellows virus (isolate FL-1) (BWV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Luteoviridae;
OC Polerovirus.
OC NCBI_TaxID=12043;
OX [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=89057523; PubMed=3194229;
RA Veit I., Lot H., Leisner M., Scheidecker D., Guillely H., Richards K.,
RA Jonard G.;
RT "Nucleotide sequence of beet western yellows virus RNA."
RT Nucleic Acids Res. 16:9917-9932(1988).
RN [2]
RP REVISIONS.
RA Veit I., Lot H., Leisner M., Scheidecker D., Guillely H., Richards K.,
RA Jonard G.;
RT Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (NMP)(N).
CC -1- SIMILARITY: 70% SIMILARITY TO POTATO LEAFROLL VIRUS ORF3.
CC AND PARTIAL SIMILARITY TO SOUTHERN BEAN MOSAIC VIRUS PUTATIVE
CC RNA-DEPENDENT RNA POLYMERASE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X13063; CAA31464.2; -
CC PIR: S01940; RVOCFL.
CC InterPro: IPR001795; Luteo_RNA_pol.
CC Pfam: PF02123; Luteo_ORF3; 1.
CC PRINTS: PR00914; LVIRSRNAPOL.
CC TRANSFERASE: RNA-directed RNA polymerase.
CC SEQUENCE 1035 AA; 115870 MW; 54642FEC88E6F66F CRC64;
SQ
Query Match 72.2%; Score 39; DB 1; Length 1035;
Best local Similarity 77.8%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 GGNSSPVF 10
DB 354 GGGSGSPV 362
-----
RESULT 10
PYRD_ARATH STANDARD; PRT; 434 AA.
AC P32746;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dihydroorotate dehydrogenase, mitochondrial precursor (EC 1.3.3.1)
DE (Dihydroorotate oxidase) (DHODHase).
OS PYRD.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
OX [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=93272056; PubMed=1303803;
RX Minet M., Dufour M.E., Lacroite F.;
RT "Complementation of Saccharomyces cerevisiae auxotrophic mutants by
RT Arabidopsis thaliana cDNAs."
RL Plant J. 2:417-422(1992).
CC -1- CATALYTIC ACTIVITY: (S)-dihydroorotate + O(2) = orotate +
CC H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- PATHWAY: Pyrimidine biosynthesis; fourth step.
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X62909; CAA44695.1; -
CC PIR: S23762; S23762.
CC InterPro: IPR001295; DHO_dh.
CC InterPro: IPR003009; FMN-enzyme.
CC Pfam: PF01180; DHODHase; 1.
CC DR TIGR0180; DHODHase; 1.
CC DR TIGR01036; pyrd_sub2; 1.
CC DR PROSITE; PS00911; DHODHASE_1; 1.
CC DR PROSITE; PS00912; DHODHASE_2; 1.
CC KW Pyrimidine biosynthesis; Oxidoreductase; Flavoprotein; FAD;
CC Transit peptide; Mitochondrion.
CC TRANSIT 1
CC CHAIN ? 434 MITOCHONDRION (POTENTIAL)
CC NP_BIND ? 401 DIHYDROOROTATE DEHYDROGENASE.
CC SEQUENCE 434 AA; 45568 MW; 186BA05F3EF49D91 CRC64;
SQ
Query Match 70.4%; Score 38; DB 1; Length 434;

```

Best Local Similarity 70.0%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGSNGSPVF 10
11111111
Db 366 TGLSGSKPLF 375

RESULT 11

HME1_MOUSE ID HME1_MOUSE STANDARD; PRT; 401 AA.
AC P09065;
DT 01-NOV-1988 (rel. 09, Created)
DT 01-FEB-1994 (rel. 28, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Homeobox protein engrailed-1 (Mo-Eu-1).
GN EN1 OR EN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93185339; PubMed=1363401;
RA Logan C., Hanks M.C., Noble-Topham S., Nallainathan D.,
RA Provart N.J., Joyner A.L.;
RT "Cloning and sequence comparison of the mouse, human, and chicken
engrailed genes reveal potential functional domains and regulatory
regions.";
RL Dev. Genet. 13:345-358(1992).
RN [2]
RP SEQUENCE OF 278-401 FROM N.A.
RX MEDLINE=88112776; PubMed=2892757;
RA Joyner A.L., Martin G.R.;
RT "En-1 and En-2, two mouse genes with sequence homology to the
Drosophila engrailed gene: expression during embryogenesis.";
RL Genes Dev. 1:29-38(1987).
RN [3]
RP SEQUENCE OF 298-401 FROM N.A.
RX MEDLINE=86079501; PubMed=2416459;
RA Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;
RT "Expression during embryogenesis of a mouse gene with sequence
homology to the Drosophila engrailed gene.";
RL Cell 43:29-37(1985).
RN [4]
RP SEQUENCE OF 321-380 FROM N.A.
RX MEDLINE=91099509; PubMed=1980115;
RA Holland P.W.H., Williams N.A.;
RT "Conservation of engrailed-like homeobox sequences during vertebrate
evolution.";
RL FEBS Lett. 277:250-252(1990).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ENGRAILED HOMEBOX FAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L12703; AAA03660.2; -
CC EMBL: Y00201; CAA68361.1; -
CC PIR: A26629; A26629.
CC PIR: A24778; A24778.
CC PIR: S13009; S13009.
CC PIR: A48423; A48423.
CC HSSP: P02836; 3HDD.
CC TRASPAC: T02016; -
CC MGI: MGI:95389; En1.
CC InterPro: IPR000747; Engrailed.

DR InterPro: IPR001356; Homeobox.
DR Pfam: PF00046; homeobox.1.
DR PRINTS: PR00026; ENGRAILED.
DR PRINTS: PR00024; HOMEBOX.
DR PRODOM: PD000010; Homeobox.1.
DR SMART: SM00389; HOX.1.
DR PROSITE: PS00027; HOMEBOX_1; 1.
DR PROSITE: PS00071; HOMEBOX_2; 1.
DR PROSITE: PS00033; ENGRAILED.1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DOMAIN 52 87 PRO-RICH.
FT DOMAIN 73 87 POLY-PRO.
FT DOMAIN 207 228 POLY-ALA.
FT DNA_BIND 312 371 HOMEBOX.
SQ SEQUENCE 401 AA; 40950 MW; 1F90210950152FAE CRC64;

Query Match Best Local Similarity 68.5%; Score 37; DB 1; Length 401;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGSNGSP 8
238 SGNMGSP 245

RESULT 12

CEL1_AGABI ID CEL1_AGABI STANDARD; PRT; 320 AA.
AC 000023;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Cellulose-growth-specific protein precursor.
GN CEL1.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RX MEDLINE=93012985; PubMed=1398098;
RA Raguz S., Yaguee E., Wood D.A., Thurston C.F.;
RT "Isolation and characterization of a cellulose-growth-specific gene
from Agaricus bisporus.";
RL Gene 119:183-190(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RX MEDLINE=94237428; PubMed=8181702;
RA Arnesilla A.L., Thurston C.F., Yaguee E.;
RT "Cell: a novel cellulose binding protein secreted by Agaricus
bisporus during growth on crystalline cellulose.";
RL FEBS Microbiol. Lett. 116:293-299(1994).
CC -1- FUNCTION: PROBABLE GLYCOSYL HYDROLASE ACTIVE ON CELLULOSE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -1- SIMILARITY: BELONGS TO FAMILY 61 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M86356; AAA53434.1; -
CC HSSP: P00725; 2CBH.
CC InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR005103; Glyco_hydro_61.
DR Pfam: PF00734; CBM_1; 1.
DR Pfam: PF03443; Glyco_hydro_61; 1.

DR SMART; SM00236; fCBD; 1.
 DR PROSITE; PS00562; CBD; FUNGAL; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
 FT SIGNAL 1 29
 FT CHAIN 30 320
 FT DOMAIN 30 261
 FT DOMAIN 262 285
 FT DOMAIN 286 320
 FT DISULFID 292 309
 FT DISULFID 303 319
 FT CARBOHYD 163 163
 SQ SEQUENCE 320 AA; 33754 MW; 60E2C80895CA2B CRC64;

Query Match
 Best Local Similarity 66.7%; Score 36; DB 1; Length 320;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSP 8
 Db 256 SGGNGSP 263

RESULT 13
 ID PLYA-COGL STANDARD; PRT; 380 AA.
 AC 000374;

DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 GN PMLA.

OS Colletotrichum gloeosporioides (Anthracnose fungus) (Glomerella
 OS clingulata).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetes Incertae sedis; Phyllosthorales; Phyllachoraceae;
 OC Glomerella.
 OC NCB1_TaxID=5457;

RP SEQUENCE FROM N.A.
 RX MEDLINE-94237480; PubMed-8181749;

RA Templeton M.D., Sharrock K.R., Bowen J.K., Crowhurst R.N.,
 RA Rikkerink E.H.;

RT "The pectin lyase-encoding gene (pnl) family from Glomerella
 RT clingulata: characterization of pnl and its expression in yeast.";
 RL Gene 142:141-146(1994).

CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectin to give
 CC oligosaccharides with terminal 4-deoxy-6-methyl-alpha-D-galact-4-
 CC enuronosyl groups.
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE LYASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; L22857; AAA21817.1; -.
 DR HSSP; 001172; 1IDJ.
 DR InterPro; IPR002022; Amb.allergen.

DR Pfam; PF00544; pec_lyase.1.
 KM Lyase; Signal; Glycoprotein.

FT SIGNAL 1 20
 FT CHAIN 21 380
 FT CARBOHYD 130 130
 SQ SEQUENCE 380 AA; 39326 MW; 3DF9A99FBA482053 CRC64;

Query Match
 Best Local Similarity 66.7%; Score 36; DB 1; Length 380;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 10
 Db 36 TGGSGATPV 45

RESULT 14
 ID Y350_HAEIN STANDARD; PRT; 425 AA.
 AC P24326;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein HI0350 (ORF3).
 GN HI0350.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OC NCB1_TaxID=727;

RP SEQUENCE FROM N.A.
 RC STRAIN=RM 7004 / Serotype B;
 RX MEDLINE-92065797; PubMed-1956282;
 RA Maskell D.J., Szabo M.J., Butler P.D., Williams A.E., Moxon E.R.;

RT "Molecular analysis of a complex locus from Haemophilus influenzae
 RT involved in phase-variable lipopolysaccharide biosynthesis.";
 RL Mol. Microbiol. 5:1013-1022(1991).

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,
 RA Kierulff A.R., Bull C.J., Tomb J.-F., Dougherty B.A., Kerrick E.F.,

OY 1 TGNSSGPVF 10
 111111
 Db 20 TGNSSGLPLF 29

RESULT 15

DEGP_RICCN STANDARD; PRT; 508 AA.
 AC 092JAL;
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR RC0166.
 OS Rickettsia conorii.
 OC Bacteria: Proteobacteria: alpha subdivision: Rickettsiales:
 Rickettsiaceae: Rickettsiae: Rickettsia.
 NCBI_TaxID=781;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: AE008583; AAL02704.1; ALT_INIT.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR Pfam: PF00595; PDZ; 2.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS0106; PDZ; 1.
 DR KW Hydrolyase; Serine protease; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 23
 FT CHAIN 1 508
 FT DOMAIN 119 284 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 286 377 CATALYTIC.
 FT DOMAIN 413 497 PDZ 1.
 FT ACT_SITE 134 134 PDZ 2.
 FT ACT_SITE 164 164 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 508 AA; 55599 MW; D2F53A690ECD0AD7 CRC64;

Query Match 66.7%; Score 36; DB 1; Length 508;
 Best Local Similarity 75.0%; Pred. No. 59;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 GNSGSPVF 10
 111111
 Db 240 GNSGSPVF 247

Search completed: December 20, 2002, 12:02:21
 Job time : 25.2632 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:11:39 ; Search time 1363.26 Seconds

(without alignments)
192.131 Million cell updates/sec

Title: US-10-008-355-26

Perfect score: 48

Sequence: 1 TCGNCSGPV 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=ylh
-Q/cgcn2.1/USPTO.spool/US10008355/runat_17122002_112337_14593/app.query.fasta_1.398
-DB=GenEmbl -OFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdi -LIST=45
-DOCAIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355.ecgn.1.1616 -runat_17122002_112337_14593 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAPP -LARGEOUTER -NEG_SCORES=0 -WAIT -LONGLOG -DEVTIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ot:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pin:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_sy:*
39: em_hgtgo_hum:*
40: em_hgtgo_mus:*
41: em_hgtgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	48	100.0	1289	1	STAGASP	D00730 S. aureus g
2	48	100.0	1558	6	E03836	E03836 DNA encodin
3	48	100.0	1586	6	E03835	E03835 DNA encodin
4	48	100.0	1634	1	SASP	Y00356 Staphylococ
5	48	100.0	3240	1	AF309515	AF309515 Staphyloc
6	48	100.0	5207	1	SWA293885	AJ293885 Staphyloc
7	48	100.0	10689	1	AE004008	AE004008 Xylella f
8	48	100.0	290150	1	AP004825	AP004825 Staphyloc
9	48	100.0	298050	1	AP003132	AP003132 Staphyloc
10	48	100.0	347235	1	AP003361	AP003361 Staphyloc
11	44	91.7	1912	9	AF008576	AF008576 Homo sapi
12	44	91.7	162609	9	AC108022	AC108022 Homo sapi
13	44	91.7	166978	2	AC108085	AC108085 Homo sapi
14	44	91.7	171279	2	AC036185	AC036185 Homo sapi
15	44	91.7	174253	2	AC068590	AC068590 Homo sapi
16	44	91.7	189014	2	AC051649	AC051649 Homo sapi
17	44	91.7	234431	2	AF311103	AF311103 Homo sapi
18	43	89.6	657	1	SEP305145	AJ0305145 Staphyloc
19	43	89.6	849	6	AX141641	AX141641 Sequence
20	43	89.6	3189	6	AF269652	AF269652 Staphyloc
21	43	89.6	3189	6	AX144972	AX144972 Sequence
22	43	89.6	62485	9	AL590093	AL590093 Human DNA
23	43	89.6	80141	9	HSEMSGAR	Y07848 Homo sapien
24	43	89.6	110959	2	AC105728	AC105728 Rattus no
25	43	89.6	122146	9	AC011736	AC011736 Homo sapi
26	43	89.6	133240	2	AC000026	AC000026 Homo sapi
27	43	89.6	133240	2	AC120820	AC120820 Rattus no
28	43	89.6	159681	2	AC036239	AC036239 Homo sapi
29	43	89.6	167960	9	AC093084	AC093084 Homo sapi
30	43	89.6	173029	9	AC002059	AC002059 Homo sapi
31	43	89.6	178273	2	AC005308	AC005308 Plasmodu
32	42	87.5	5640	1	AB011418	AB011418 Alteromon
33	42	87.5	11136	1	AE000862	AE000862 Methanoba
34	42	87.5	20993	10	AB043785	AB043785 Mus muscu
35	42	87.5	43320	9	AC005784	AC005784 Homo sapi
36	42	87.5	74073	2	AC111139	AC111139 Mus muscu
37	42	87.5	90923	2	AL159157	AL159157 Homo sapi
38	42	87.5	133925	9	AC011444	AC011444 Homo sapi
39	42	87.5	145905	9	AC087711	AC087711 Homo sapi
40	42	87.5	148228	2	AC051659	AC051659 Homo sapi
41	42	87.5	167320	2	AF298853	AF298853 Homo sapi
42	42	87.5	170340	2	AC092703	AC092703 Homo sapi
43	42	87.5	186136	2	AC110433	AC110433 Rattus no
44	42	87.5	342800	1	AP003598	AP003598 Nostoc sp
45	41	85.4	1085	1	EFSPREG	Z12296 E. faecalis

RESULT 1

ALIGNMENTS

STAGASP STAGASP 1289 bp DNA linear BCT 01-FEB-2000
LOCUS S. aureus glutamic acid specific protease (EC 3.4.21.19) gene.
DEFINITION D00730
ACCESSION D00730.1 GI:216970
VERSION V8-like protease; glutamic acid specific protease.
KEYWORDS S. aureus (strain ATCC12600) genomic DNA.
SOURCE Staphylococcus aureus
ORGANISM Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1 (bases 1 to 1289)
AUTHORS Yoshikawa, K., Tsuzuki, H., Fujiwara, T., Nakamura, E., Yamoto, H.,
Matsumoto, K., Shin, M., Yoshida, N. and Teraoka, H.
TITLE Purification, characterization and gene cloning of a novel glutamic
acid-specific endopeptidase from Staphylococcus aureus ATCC 12600
JOURNAL Biochim. Biophys. Acta 1121 (1-2), 221-228 (1992)
MEDLINE 9287954
COMMENT Submitted (30-OCT-1990) to DDBJ by:
Etsuo Nakamura
Shionogi Research Laboratories
SHIONOGI & Co., Ltd
Fukushima-ku, Osaka 553
Japan
Phone: 06-458-5861 x571
Fax: 06-458-0987.
FEATURES
source
1. .1289
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
55. .1128
/EC_number="3.4.21.19"
/codon_start=1
/transl_table=11
/product="glutamic acid specific protease prepropeptide"
/protein_id="BAA00630.1"
/db_xref="GI:216971"
/translation="MKGKFLKYSILFVATLTITATVSSPANALSSKANDHPQOTQ
DKQTPKIQKGNLKPLEDKERANVILPNNDRHQITDTNGHYAVVYIQVEAPGTE
IASGVVSGKDTLLINKHVVADATGDPHALKAPSAINQNPNGGFTAEQITKSGEG
DLAIVKSPNEONKHIGEYVKPATMSNAETGVNONITVGYPGKRPVATMESGKI
TYLKGEMOYDLSTTGNGSGSPFENKKNENVIGIHHGCVNPGFNFINEVNRNLKO
NIEDINPANDHPNNPDNPNNDPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPNPN
DQNNPNPNNDNNNDNNNDPNDA"
misc_feature
919. .1089
/note="(Prokaryon) sequence"
BASE COUNT 484 a 251 c 202 g 352 t
ORIGIN
Alignment Scores:
Pred. No.: 187 Length: 1289
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-10-008-355-26 (1-9) x STAGASP (1-1289)
QY 1 ThrGlyGlyAsnSerGlySerProval 9
Db 751 ACTGCTGTAACACGCTCACCCTGTA 777
RESULT 2
E03836 E03836 1558 bp DNA linear PAT 29-SEP-1997
LOCUS DNA encoding V8 protease.
ACCESSION E03836
VERSION E03835.1 GI:2172050
KEYWORDS JP 1992211370-A/2.
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus

Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1 (bases 1 to 1558)
AUTHORS Nakamura, E., Tamaki, M., Teraoka, H., Matsumoto, K., Shin, M.,
Fujiwara, K., Tsuzuki, H., Yoshida, N. and Kakudou, S.
TITLE NEW PROTEASE
JOURNAL Patent: JP 1992211370-A 2 03-AUG-1992;
SHIONOGI & CO LTD
COMMENT Staphylococcus aureus
PN JP 1992211370-A/2
PD 03-AUG-1992
PF 19-FEB-1991 JP 1991024633
PR 20-FEB-1990 JP 90P 40398
PI NAKAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI MATSUMOTO
KOICHI,
PI SHIN MASARU, FUJIMARA KOJI, TSUZUKI HIROSHIGE, YOSHIDA NOBUO,
PI KAKUDOU SHINJI
PC C12N9/52,C12N15/57,(C12N15/57,C12R1.445);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
CC *source: strain=V8;
CC *source: clone=PAM82SIGV8;
CC feature is identified by similarity;
FH key Location/Qualifiers
FH
FT CDS 354. .1364
FT sig_peptide 354. .557
FT mat_peptide 558. .1361
FT Location/Qualifiers
1. .1558
/organism="Staphylococcus aureus"
/db_xref="taxon:1280"
BASE COUNT 579 a 282 c 221 g 476 t
ORIGIN
Alignment Scores:
Pred. No.: 212 Length: 1558
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-008-355-26 (1-9) x E03836 (1-1558)
QY 1 ThrGlyGlyAsnSerGlySerProval 9
Db 1050 ACTGCTGTAATCAGCTCACCCTGTA 1076
RESULT 3
E03835 E03835 1586 bp DNA linear PAT 29-SEP-1997
LOCUS DNA encoding V8-like protease.
ACCESSION E03835
VERSION E03835.1 GI:2172049
KEYWORDS JP 1992211370-A/1.
SOURCE Staphylococcus aureus.
ORGANISM Staphylococcus aureus.
REFERENCE 1 (bases 1 to 1586)
AUTHORS Nakamura, E., Tamaki, M., Teraoka, H., Matsumoto, K., Shin, M.,
Fujiwara, K., Tsuzuki, H., Yoshida, N. and Kakudou, S.
TITLE NEW PROTEASE
JOURNAL Patent: JP 1992211370-A 1 03-AUG-1992;
SHIONOGI & CO LTD
COMMENT Staphylococcus aureus
PN JP 1992211370-A/1
PD 03-AUG-1992
PF 19-FEB-1991 JP 1991024633
PR 20-FEB-1990 JP 90P 40398

PI	NAMAMURA ETSUO, TAMAKI MIKIO, TERAOKA HIROSHI, PI	MATSUMOTO KOICHI,
PI	SHIN MASARU, FUJIMARA KOIJI, TSUZUKI HIROSHIE,	YOSHIDA NOBUO,
PI	KAKUDOU SHINDO	
PC	C12N9/52, C12N15/57, (C12N15/57, C12R1:445);	
CC	strandedness: Double;	
CC	topology: Linear;	
CC	hypothetical: No;	
CC	anti-sense: No;	
CC	*source: strain=ATCC12600;	
CC	*source: clone=PAM82SIGV8;	
CC	Feature is identified by experimental;	
FH	Key	Location/Qualifiers
FH		
FT	CDS	352..1425
FT		
FT	sig_peptide	352..555
FT	mat_peptide	556..1422
FT		
FEATURES		
source	Location/Qualifiers	
	1..1586	
	/organism="Staphylococcus aureus"	
	/db_xref="taxon:1280"	
BASE COUNT	589 a	302 c 219 g 476 t
ORIGIN		

Alignment Scores:	
Pred. No.:	215
Score:	48.00
Percent Similarity:	100.00
Best Local Similarity:	100.00
Query Match:	100.00
DB:	6
Length:	158
Matches:	9
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-008-355-26 (1-9) x E03835 (1-1586)

	QY
	1 ThrGlyGlyAsnSerGlySerProVal 9
Db	1048 ACTGTCTAAGTCAAGTTCAACCTTA 1074

RESULT 4		
SASP		
LOCUS	SASP	1634 bp DNA linear
DEFINITION	Staphylococcus aureus V8 serine protease gene.	BCT 27-MAR-1995
ACCESSION	Y00356	
VERSION	Y00356.1	GI:46686
KEYWORDS	serine protease.	
SOURCE	Staphylococcus aureus.	
ORGANISM	Staphylococcus aureus.	
REFERENCE	Bacteria; Firmicutes; Bacillales; Staphylococcus.	
AUTHORS	1 (bases 1 to 1634)	
TITLE	Gray, G.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (16-SEP-1987) Gray G., Genencor, Inc., 180 Kimball Way,	
AUTHORS	South San Francisco, CA 94080, USA	
TITLE	2 (bases 1 to 1634)	
JOURNAL	Carmona, C. and Gray, G. L.	
REFERENCE	Nucleotide sequence of the serine protease gene of Staphylococcus	
AUTHORS	aureus, strain V8	
TITLE	Nucleic Acids Res. 15 (16), 6757 (1987)	
JOURNAL		
MEDLINE	8716953	
PUBMED	3306605	
FEATURES		
SOURCE	Location/Qualifiers	
	1..1634	

```

354 . . 1.364
/nc="preproenzyme (AA -68 to 268)"
/codon_start=1
/transl_table=1
/protein_id="CMA68434.1"

```

```

mat_peptide
538.1361
/product="mature serine protease (AA 1-268)"

misc_feature
354..557
/note="prepro region (AA -68 to -1)"

misc_feature
1218..1325
/note="twelvefold repeated tripeptide (Pro-Asp-Asn)"

BASE COUNT
617 a 296 c 231 g 490 t

ORIGIN
```

Alignment Scores:	
Pred. No.:	219
Score:	48.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	100.00%
DB:	1
Length:	163
Matches:	9
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-10-008-355-26 (1-9) x SASP (1-1634)

QY	1	Th	rg	ly	as	n	s	e	r	g	l	y	s	e	r	p	r	o	v	a	l	9
Db	1050	A	C	T	G	T	G	T	A	T	T	C	A	G	G	T	T	C	A	C	C	T
		G	T	A	T	T	C	A	G	G	T	T	C	A	C	C	T	G	T	A	1076	

```

RESULT 5
AF309515
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

```

AF309515 3240 bp DNA linear BCT 18-JAN-2001

Staphylococcus aureus serine protease operon, complete sequence.

AF309515

AF309515.1 GI:12025237

Staphylococcus aureus.

Staphylococcus aureus

Bacteria; Firmicutes; Bacillales; Staphylococcus.

1 (bases 1 to 3240)

Rice, R., Perilla, R., Bast, D., de Azavedo, J. and McGavin, M.J.

Description of staphylococcus serine protease (ssp) operon in staphylococcus aureus and nonpolar inactivation of sspa-encoded serine protease

Infect. Immun. 69 (1), 159-165 (2001)

20569178

11119502

2 (bases 1 to 3240)

Rice, R.C. and McGavin, M.J.

Direct Submission

Submitted (28-SEP-2000) Laboratory Medicine and Pathobiology, University of Toronto, Sunnybrook and WCHSC, 2075 Bayview Avenue, Toronto, ON M4N 3M5, Canada

Location/Qualifiers

1..3240

```

gene      354. .2994
           /gene="serine protease operon"
gene      354. .1364
           /gene="sspa"
           354. .1364
CDS       /gene="sspa"
           /note="serine protease; V8 protease; glutamyl
           endopeptidase"
           /codon_start=1
           /transl_table=11
           /product="sspa"
           /protein_id="AAC45843.1"
           /db_xref="GI:12025238"

```

/translation="MKGRFLKVSLEFVALITLITATLVSSPANAALLSKAMDNHPQOTQS
SKOQTPKIQKGNLKPLEQREHANVILPNNNDHQIITDTNGHAPVTVIOVAPRGTF
IASGVVYKDTLLTNKHYVDATGDPHALKAPSAITNODNNGYNGGATFQIOTKRSSEG
DLAIKVESPNQONKHIGEVYKPAATSNMAETOVNONITVTGYPGDKPVATIMESGKI
TYLKGAHOYDLSSTTGSGNSGSPVFNKKNVITIHGGVFNENGAVFIMENRNPFLKO
NIIEDHIFANDDQPNNDPNNDPNNDPNNDPNNDPNNDPNNDPNNDPNNDPNNDPNND
AA"
1446..2627
/gene="sspb"
1446..2627
/gene="sspb"
/note="sspb: preproprotein"
/codon_start=1
/transl_table=11
/product="cysteine protease precursor"
/protein_id="AAG45844.1"
/db_xref="GI:12025239"
/translation="MNSCKSRVFNIIISIMVSMILSLGAFAANNKRAKADSHSKOLE
INVKSDKVPQKVDLAQOQFAGYAKALDQSNAKGKELGEAFKIKNGEDNSY
YVINKDKIIVYTLTLSPKNDLNSKEDMNSVKSISNFIAKDLDQIKDKNSITVILT
DKRGFEEDGKRVLYKATPLPGNVKESAKTYSAKIKQFLKNTVTPKRVENENAIQ
EDVOYKENTLAKNFKIREQOFDMSWCGESMALLNKTNTDTYNAHDIMRTIYPEYSE
QDLPGCATFPNQNIETGKSQGRDITHQSGVSPSTEQVDLTQDNVGMILAQSVQPNP
DHLGHALAIVGNAKINDQEKLIYNNPMDTELSIDADSSLHLSPFRDYMVYGSWIG
Y"
1446..1553
/gene="sspb"
2103..2624
/gene="sspb"
/product="cysteine protease"
1554..2624
/gene="sspb"
/note="encodes the cysteine protease proprotein"
2655..2994
/gene="sspb"
2665..2994
/gene="sspc"
2665..2994
/gene="sspc"
/note="unknown; possesses no signal sequence; predicted
cytoplasmic location"
/codon_start=1
/transl_table=11
/product="sspc"
/protein_id="AAG45845.1"
/db_xref="GI:12025240"
/translation="MYOLOFINLVYDVTTLTLEQTNINLFLIGNNSNHLOKSGICIRH
GDDTSHNOYHILFIDTAHORIKFSSIDNEIITIIYIDYDTQIHILMQTSKQIGISRP
IYERLY"
BASE COUNT 1254 a 547 c 461 g 978 t
ORIGIN
Alignment Scores:
Pred. No.: 347 Length: 3240
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-10-008-355-26 (1-9) x AF309515 (1-3240)
Oy 1 ThrGlyGlyAnsSerGlySerProval 9
|||||
Db 1050 ACTGTTGTAATTCAGTTCACCTGTA 1076
RESULT 6
SMA293885 5207 bp DNA linear BCF 12-AUG-2002
LOCUS Staphylococcus warneri prom gene, proc gene, prod gene, spw1 gene
DEFINITION and msw1 gene.
ACCESSION AJ293885
VERSION AJ293885.2 GI:22218025
KEYWORDS Cysteine proteinase; glutamyl endopeptidase; msw1 gene; proc gene;
prom gene; prod gene; spw1 gene.

SOURCE Staphylococcus warneri.
ORGANISM Staphylococcus warneri
REFERENCE Bacteria: Firmicutes; Bacillales; Staphylococcus.
AUTHORS 1
Yokoi, K., Kakikawa, M., Kimoto, H., Watanabe, K., Yasukawa, H.,
Yamakawa, A., Takeo, A. and Kodaira, K. I.
TITLE Genetic and biochemical characterization of glutamyl endopeptidase
of Staphylococcus warneri M
JOURNAL Gene 281 (1-2), 115-122 (2001)
MEDLINE 21623048
REFERENCE 2
Kakikawa, M.
AUTHORS Direct Submission
TITLE Submitted (30-AUG-2000) Kakikawa M., Molecular Biology Group,
JOURNAL Toyama University, 3190 Gofuku, Toyama, 930-8555, JAPAN
REMARK revised by author [08-AUG-2002]
COMMENT On Aug 13, 2002 this sequence version replaced gi:9968802.
FEATURES
source
gene
CDS
gene
CDS
mat_peptide
gene
CDS
mat_peptide
1..5207
/organism="Staphylococcus warneri"
/db_xref="taxon:1292"
complement(232..708)
/gene="msw1"
complement(232..708)
/gene="msw1"
/codon_start=1
/transl_table=11
/protein_id="CAD43735.1"
/db_xref="GI:22218026"
/translation="MEIYVAGCGCLMGVQAFKRTIPGIVSTEGRANGITSNLNGEYD
GYAEVCKLTFDPDSLIRINMSYLFEDVDPYSLNQGQDQIGKRYTGLYSENTKHLNE
AQTFINEREDRRAHIVVELYPLSYKSAPEHQDRLENYEDHILCHIRPMHMLNKKY"
991..2148
/gene="spw1"
991..2148
/gene="spw1"
/codon_start=1
/transl_table=11
/protein_id="CAD43736.1"
/db_xref="GI:22218027"
/translation="MTHSIKYLIGATTTTSLALFLTSHHEASATENNQNDQGNHO
SNGNAITQANQTDSTYNNHDENGPIAVDNNIDNNHNOGDAIVSDNVDAATFNHDE
ACNPPLMTAVNSVDFNHDGNGPITAEENNIDSDFNHDGNGPIAVDNNIDNNHNOG
GDAIVSDNVDAVSNFHNDEGNGPLMTANSVDSDFNHDGNGPIASDNNIDNNHNOG
DAIVSDNNIDNNHNOGNGEALYANKNADYGFPHHDVAGCPITAHNNQAMISRNQOSHDT
PIASDNNIDNNHNOGNGEALYANKNADYGFPHHDVAGCPITAHNNQAMISRNQOSHDT
TTPQSAITTHHOSNSLPATGQDRTAPYIGFVITLVLAFAFLTKRSSAKK"
2432..3382
/gene="prom"
2432..3382
/gene="prom"
/codon_start=1
/transl_table=11
/product="glutamyl endopeptidase"
/protein_id="CAC06168.1"
/db_xref="GI:9968803"
/db_xref="SPTREMBL:Q9FRG1"
/translation="MKVKEFTASLLILATLTSATLINPAHAFTSTSTDNHOOTDSOQ
QKTPKIDGNNNVPEKKERANVILPNDNRHDIINDTTLGHTAPVTFVVOVSREGTFIA
SGVVVGKDKLLTNKHYVDATGHPRALAFAPSAVQNNYVNGGFTGEOITKPGNADL
AIVKFSPPNNQONHIGEVYTPATLSNADTSQNPITVYGYPGDKPLATIMESRGKITO
IOGEDMHYDLSSTTGSGNSGSPVFNKKNVITIHGGVFNENGAVFIMENRNPFLKONI
EDIHFSNSDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNNDNNND
2620..3378
/gene="prom"
/product="glutamyl endopeptidase"
/evidence=experimental
3471..4661
/gene="proc"
3471..4661
/gene="proc"
/codon_start=1
/transl_table=11


```
/product="cysteine proteinase"
/protein_id="CAD43737.1"
/db_xref="GI:22218028"
/translation="MHSHYKYSFKLSTLFAILLLSIAFTDLNKSVAACKNSQIE
INVKNDKVPQKYNLAACOYKCYAQLDKKNSQSDYDGLGAFKIKYKNGEDNSY
YVLVDGLKLTITLISPKNEADLKOSKREANYSVKSINIAADLEIKKHNTITLT
DKRGYFEENGVRLVKAPLPTNINKNSQMSHSLKQOLKASEPKINENQAAE
ANQDNQOVYENTLKNEKIREQDFNSWCAGFSMAALLNATKTKDKYNADIMRTLYPN
VSEELPEPSTPHQMIKYGNSQGRIOHQSGMPSPYNIDOLTCKOKGAMIIASHVEN
NNPDHLGHMAVNCNAKINNNDKLIFFNPMDTDSIQDADSNLHLSNRPDTYDDT
MIGY"
/gene="5024
/feature="prod"
/feature="5024
/feature="5024
/feature="prod"
/codon_start=1
/transl_table=11
/protein_id="CAD43738.1"
/db_xref="GI:22218029"
/translation="MYOLHFHINDALTLFSKODTILHFGINWINSQKSIIRT
GYDTNHNQYQIIQIDTEHORIKLTSEDPQLMWILDYEDTNHIFLOTYSKNSYGRS
IYKEK"
BASE COUNT 1937 a 988 c 751 g 1531 t
ORIGIN
Alignment Scores:
Pred. No.: 476 length: 5207
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-10-008-355-26 (1-9) x SMA293885 (1-5207)
QY 1 ThGlyGlyAsnSerGlySerProval 9
Db 3122 ACTGTGTGATTCAGGTTCACCTGTA 3148
RESULT 7
LOCUS AE004008 10689 bp DNA linear BCT 15-JUN-2001
DEFINITION Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.
ACCESSION AE004008 AE003849
VERSION AE004008.1 GI:9106961
KEYWORDS
SOURCE Xylella fastidiosa 9a5c.
ORGANISM Xylella fastidiosa 9a5c
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.
1 (bases 1 to 10689)
REFERENCE
AUTHORS Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M., Araya,J.E., Bala,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.,
Bueno,M.R., Camargo,A.A., Costa,L.E., Carraro,D.M., Carrier,H.,
Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M.,
Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H.,
Facinanci,A.P., Ferreira,A.J., Ferreira,V.C., Ferro,J.A.,
Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R.,
Gardier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A.,
Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P.,
and Marino,C.L.
TITLE The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis
NATURE 406 (6792), 151-157 (2000)
JOURNAL
MEDLINE 20365717
PUBMED 10910347
AUTHORS 2 (bases 1 to 10689)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
```

```
Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorri,H., Facinanci,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
Frohme,M., Furlan,L.R., Gardier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Klieger,J.E., Kuramae,E.E., Laigret,F.,
Lambis,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.F.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
Tsubako,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S.,
Vettore,A.L., Zaio,M.A., Zatz,M., Meidanis,D. and Setubal,J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil
FEATURES
source
LOCATION/Qualifiers
1..10689
/organism="Xylella fastidiosa 9a5c"
/db_xref="taxon:160492"
/clone="9a5c"
/complement(119..367)
/gene="Xf1871"
/complement(119..367)
/gene="Xf1871"
/feature="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AF84677.1"
/db_xref="GI:9106962"
/translation="MTMTLEKDEKTEIIDEORIQIAKLSEFAKIQAEIHEVNAHQKL
MKETLKTETESKWPVVGSSSLIAAGATATLFIKIFH"
/complement(430..618)
/gene="Xf1872"
/complement(430..618)
/gene="Xf1872"
/feature="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AF84678.1"
/db_xref="GI:9106963"
/translation="MMEFAARAPOGGLAASDPQRTDAPKKTGLKKRKYPTGASASI
TVIRHVLTRQOKRKASP"
941..1288
/gene="Xf1873"
941..1288
/gene="Xf1873"
/feature="similar to GI1479340 (percent identity: 36 %/query
alignment coverage: 87.8 %/subject alignment coverage:
90.2 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AF84679.1"
/db_xref="GI:9106964"
/translation="MSKRNIPQIILGNMGAGQFINGKLVTPERTITEEMQWLSTL
CTIARQMMEAEARATAPOGGLAASDPQRTDAPKKTGLKKRKYPTGASASI
VLIHQQRKKRASP"
941..1288
```

gene 1351..1662
/gene="XF1874"
CDS 1351..1662
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84680.1"
/db_xref="GI:9106965"
/translation="MTMTKEDKTEIDQKIQIAKLISETAKIQAEHEVYTHQEM
NAHINEMARTQEMNARTQKLIETQKVKESQWYFVVASGLLAGATAATLTKLF
H"
complement(1839..2918)
/gene="XF1875"
CDS complement(1839..2918)
/gene="XF1875"
/note="similar to SPIP03626 (percent identity: 36 %/query
alignment coverage: 88.0 %/subject alignment coverage:
105.0 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84681.1"
/db_xref="GI:9106966"
/translation="MLNLTGAPGNGKTLAYDMLRQIEIDKSLVKSGAVRSEFYTD
IGGFYEAVRRLTGYVQSAPEDMRTTPGSGYIVDEARMPAGRGSDDEPRVCDI
DTHRHGGIDLMFVTQMPKIHHLRLVEHVLHAKMLQTAGLRMSRADDDPDI
HOREKAEVEWKFEPKDRYALYASSTLHTVSHKFRIPKYMALSVCTCSVIGLVWY
YSPALHSEASVAGASQASLRAPALSSRSLSGRTYAVLETFSAPTLSCVC
SSDRSCRCNTDGYQIDMSVGECCRLRLAPLPFNHYHAVYVSSSSAAVSSPSSPS
SSSVSLSSVSSPILGSSSVPSH"
complement(2918..3202)
/gene="XF1876"
CDS complement(2918..3202)
/gene="XF1876"
/note="similar to SPIP03625 (percent identity: 36 %/query
alignment coverage: 85.1 %/subject alignment coverage:
86.0 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phage-related protein"
/protein_id="AAF84682.1"
/db_xref="GI:9106967"
/translation="MPSWLAMLRNLGOTVARGVAGLAIVTTPLPIVTSALNL
IVSKSGISADVNLNLMLMGFEALSIISAMLTRIALQSLHGVKATP"
complement(3195..4412)
/gene="XF1877"
CDS complement(3195..4412)
/gene="XF1877"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84683.1"
/db_xref="GI:9106968"
/translation="MCSRSFLEFYISALGLTIVISPFCLFAQNPVVEVPTSVRY
SSVSDGVSTATFEARSTALNGVYVVPDISASTIGSLAKYAVRGMAFYVYS
MISGLNGAGWIDELTHEVMGSPALKEIPGVVAMMORPGDGHVYFVPOGLIAP
INAYOSLSPOPIYVSSGSSVSSSRMLXHLGGGEVYHLSKTDSPDXSSGILP
SYILDDLGQVRSDDSTYNAVLTDSQTAVALTPEITYSALKLRSLDELTAKASAP
DOOPSSGASPPSSSGTAVPSCFASVAVCDFTDWKSDDELKRPPLVPDPVYVDKL
PQAKTWSGLGEGACSPPTTPIEFSGYKTSVEISYQPCDEPALMRPVIVIAITIA
AVIAGFRGVKXV"
complement(4481..4702)
/gene="XF1878"
CDS complement(4481..4702)
/gene="XF1878"
/note="hypothetical protein; identified by sequence

similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84684.1"
/db_xref="GI:9106969"
/translation="MKSYPDRCIAFFSLSTSSLVFADADAGAFDAGAAVTALGAIS
GAVALIGAKLAPPAISVGMKWLKAIFG"
complement(4724..4924)
/gene="XF1879"
CDS complement(4724..4924)
/gene="XF1879"
complement(4724..4924)
/gene="XF1879"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84685.1"
/db_xref="GI:9106970"
/translation="MAQQLMSLYCTQYDVEARTCSQAAMWPPSLPISYBDVRIIL
PIHVMCFVAMGEHFLFTVVRD"
complement(4931..5236)
gene
Alignment Scores:
Pred. No.: 771 Length: 10689
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-008-355-26 (1-9) x AE004008 (1-10689)
QY 1 ThrGlyGlyAsnSerGlySerProval 9
Db 8561 ACTGGCGGCAACTCGGATCGCAATC 8535
RESULT 8
AP004825 290150 bp DNA linear BCT 02-JUL-2002
AP004825/c
Staphylococcus aureus subsp. aureus MM2 DNA, complete genome,
LOCUS strain:MM2, section 4/10.
DEFINITION
AP004825 BAO000033
ACCESSION
AP004825.1 GI:21203989
KEYWORDS
SOURCE
ORGANISM
Staphylococcus aureus subsp. aureus MM2 (strain:MM2) DNA.
Staphylococcus aureus subsp. aureus MM2
REFERENCE
Bacteria; Firmicutes; Bacillales; Staphylococcus.
1
Baba,T., Takeuchi,F., Kuroda,M., Yuzawa,H., Aoki,K., Oguchi,A.,
Nagai,Y., Iwama,N., Asano,K., Naimi,T., Kuroda,H., Cui,L.,
Yamamoto,K. and Hiramatsu,K.
Genome and virulence determinants of high virulence
community-acquired MRSA
Lancet 359 (9320), 1819-1827 (2002)
JOURNAL MEDLINE 22040717
PUBMED 12044378
REFERENCE
2 (bases 1 to 290150)
Aoki,K., Oguchi,A., Nagai,Y., Asano,K., Iwama,N., Baba,T.,
Kuroda,M., Hiramatsu,K. and Kikuchi,H.
Direct Submision
Submitted (06-MAR-2002) Akio Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center, 26home 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/
tel:81-3-3481-8423, Fax:81-3-3481-8424)
Location/Qualifiers
1. 290150
/organism="Staphylococcus aureus subsp. aureus MM2"
/strain="MM2"
/db_xref="taxon:196620"
285. 1493
/gene="MW0823"

285. 1493
/gene="MW0823"
/note="ORFID:MW0823
hypothetical protein, similar to NADH dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB94688.1"
/db_xref="GI:21203990"
/transl_table="MA0DRKKVVLVAGYAGLQTVFKLQKAISTEAEAILINKNEYH
YEATMLHSAAGSLNVEDVILPVESYLKDKVNFVQAEVRKIDDAKRVETNGIYFE
DILVAGALFVESETGIEGMRKHAQFENVTIARLSLHIEDKRFANVAASKEKDNDIS
ILVAGAGTGEPIGLTDLDRIPRELCSYGDONQKVTCTCEAPAKMLPMSSEBVLNHA
VSYLEDREVVERKATPIPVACNEKGFVEVDGEKQALNAGSVMAVAGVGLHDEESE
GKRRRIATDTLNGINDNIFVIGDSSAFIPAGEEERLPPTAOLAQOGEVAKNKR
RILNGESTEEDEEYVDRGTVCISGSHDGVGMVFDKPIAGKRAAFMKKVIDIRAVFKIGG
IGLAFKKKGF"
1624. 3099
/gene="ampa"
1624. 3099
/gene="ampa"
/note="ORFID:MW0824"
/codon_start=1
/transl_table=11
/product="probable cytosol aminopeptidase"
/protein_id="BAB94689.1"
/db_xref="GI:21203991"
/transl_table="MNFELNNTLSNEINTLLIGIPHEHLNQLERISFNHIDITESLERL
KHOHIGSKVSKYITTAADVDODQTVRLITVGLNLKTRSYQODMLKIGHLFOYKSHSH
IEDPLDLSITSKYDQDLSVDLMACGISOSEKATVEPDHYSKKAPKPTNLNLSSEL
IELPIHIGSISIGOSINLARDFSNMPRVLTPOFAEDYVNHENKNTVKKYDKYDPL
VSEGFGLLQAGGSKHKPRVLVTITVNGKDEAPIALVCGITVDSGGYSIKTKNKM
ATMKPDMCGAANVVGIIIEAASRIQLPVNIVGLVCAENMIAEASMKPDVFTALSGSTP
VEVNMTDGEGLVLADAVEFYANQYQSYIMGVLFATLGAALVAILDGDKAFAESKVIY
LNDLIOISEEDVEMFELPITATERASIKSHSDIADLVNHTNGQKALFAASFVTHFSG
QSPHHPFIADPATITNKASVNGPRGPGFMFTPIVQMLKQO"
3510. 4826
/gene="MW0825"
3510. 4826
/gene="MW0825"
/note="ORFID:MW0825"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94690.1"
/db_xref="GI:21203992"
/transl_table="MINAVVAVATIMVLCRLNVVISLFSIALVGLISQMSIEVY
INVGKNIVDAEVALSRLLGCFRALISYSGITDYLVGKTIKAIHENSRRMSVKKV
VTIILALAMSINSLIPVIAATPIVPIPLSLFNDLKIDRLIGLILGFGCEPY
VLLPGEFQIQIILQSGFAKAMPHEFNMLMKMLPLPSMOYIGLLIGIGVYKRPPE
YETELRISQSDNVELKRPYILIVITVAILAEVLPOTFDSMIFGALVLFISRAVN
WEELDAKVESEIKIMAYIGVILFANGAGVMNATGIDELVKTLSITGDNKFSII
MSVYLVTVTGIGSGSFATPIIILASFLPFEASIGLDMMLIALIGTASALSGDSGA
SDSLTIGPAGTAVNGQGHMIDRCVPRNLFINIDMLIFGTIANVY"
4845. 5219
/gene="MW0826"
4845. 5219
/gene="MW0826"
/note="ORFID:MW0826"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB94691.1"
/db_xref="GI:21203993"
/transl_table="MTHLETFEMSIDHOEDGLVVISMPYTDKQKPGYLHGASATK
LGEFASGSGSANLIDTYKTFPIGLLEMANNAHHSKADGHVATATETIHQGSSTHYMDIK
IKNDSEQLITVMRGVAILKPLK"
/gene="MW0827"
/gene="MW0827"
complement(5275. 6429)
/gene="MW0827"
/note="ORFID:MW0827"
/codon_start=1
/transl_table=11

/product="hypothetical protein"
 /protein_id="BAB94692.1"
 /db_xref="gi:21203994"
 /translation="MHMTIIIGGGIOGFPAIOAKLLSSGLTDRLTIIIDPHEPFOORFNS
 FENRIEMVLIASPIYVHHVHPQPHLKOPKAOHOYTNNAPFGYORPELTMMHCHAHAS
 KOYQEDCLVQGLVOTLDRKQEDKRNHIKEDSOITITTCQVYIAISTNIPRPDILKQ
 QUNVNIIFERKHQVVDKTDTHIVSSGITYAAHLAKLNLNHDNRKILHLMLKNDIEIDHF
 DADPQNPKNMSTFKSPKSMRERNAIVORRHRNSMPHELKYLKFKHINGRNLRKH
 TPITQISGLITENNSNPVYOQIIVATGFEEDFMSOPLIKOLIONYAPINECNYPIV
 SKLEWMINLFPAGCFADLELPGPRNVMGSKRAERIEQAFLKQGYSA"
 complement(6678. .7034)
 /gene="mhgc"
 /complement(6678. .7034)
 /gene="mhgc"
 /note="ORFID:MM0828"
 /codon_start=1
 /transl_table=11
 /product="Na+/H+ antiporter subunit"
 /protein_id="BAB94693.1"
 /db_xref="gi:21203995"
 /translation="MIKIIITSLALIFVYIIGALISALAAIGLRLBEDVSRANHAGAA
 SLTGAMSLLEGFELFEIATQGFVNMQLIVAILFVLITGPLSSHHMKAAYNAIKIPYRK
 KTKVDEISEDKDKPLK"
 complement(7012. .7305)
 /gene="mhpf"
 /complement(7012. .7305)
 /gene="mhpf"
 /note="ORFID:MM0829"
 /codon_start=1
 /transl_table=11
 /product="Na+/H+ antiporter subunit"
 /protein_id="BAB94694.1"
 /db_xref="gi:21203996"
 /translation="MNHNVITIVALLITIVISMLMLIRVYLGSPLADRVVALDAIGQ
 LMAVIALSILNITKYMIVYIMHIGIILAFLTAVFSKFMKGKVIETHDQNTD"
 complement(7305. .7784)
 /gene="mhhe"
 /complement(7305. .7784)
 /gene="mhhe"
 /note="ORFID:MM0830"
 /codon_start=1
 /transl_table=11
 /product="Na+/H+ antiporter subunit"
 /protein_id="BAB94695.1"
 /db_xref="gi:21203997"
 /translation="NAVOLVLFITIAVNFLEVTSTYNNVLCGFTFGLVLYILHRY
 LPRGRYVITLKRILKILVIFLEILKRAFDVLKIIIPKINERGFVYHTDLKKMQQ
 IVLNLTLTLPQGVIVGSDDRKRIYIHAIIDFSTKBOEVSIKTSLEKIVREVEL"
 complement(7786. .9282)
 /gene="mhhd"
 /complement(7786. .9282)
 /gene="mhhd"
 /note="ORFID:MM0831"
 /codon_start=1
 /transl_table=11
 /product="Na+/H+ antiporter subunit"
 /protein_id="BAB94696.1"
 /db_xref="gi:21203998"
 /translation="WIESNNMLVLTIVIPVITAILLVEIGKRPIIKRYVALGSLTLV
 AAILMANVYKHGPIRELVGSMKAPYSIVPLDIFESALLITSIIITAVILVSYQPI
 GIERRYYVVEFELFMELIGITGAFIDGIPNMLPVFEVFLMSYELLIGSTKLOQEE
 TKLYAVLVNWSSEFPVGMVALIYVSGVGLNLADLSNNKASADSGLNVNIPFLFLE
 VTAAGVAGPMPVMPSPSAVYVAPPIPIIAFSCALLITKQGYVAIAPTLSPFSDNSFHS
 YVILFLALLTITIFGCYGVAVANIKITILYNVMTAVGVYILGVAMHRESGNICATYTT
 LHMILVKALFLLIGIMIKITGTADLRQFGGLIKRYPVLWSFPTALSLAGITPPLSG
 FYGKFFIVQSEFERGEVLSGVIVLLSSLVAVYSVIRIFLGFGQPGYDLPNNKVDKQ
 YLTPTAIAVAVTVIYVGLSADYLYPMVAKAGAEFYENPSTYKAVILGSK"
 complement(9275. .9616)
 /gene="mhhc"
 /complement(9275. .9616)
 /gene="mhhc"
 /note="ORFID:MM0832"
 /codon_start=1

```
/transl_table=11
/product="Na+/H+ antiporter subunit"
/protein_id="BAB94697.1"
/db_xref="GI:21203999"
/translation="MEIIMIFVSGITFAISVYLKSLRIITVMGTTLTHANFLI
TMGGLKHCIVPIYEANVKSYPDIPQALITRAIVAFATFAFLVAFRTKELGTDN
VESMKGVPEDD"
complement(9616..10044)
/gene="mnhg"
complement(9616..10044)
/gene="mnhg"
/feature="ORFID:M0833"
/codon_start=1
/transl_table=11
/product="Na+/H+ antiporter subunit"
/protein_id="BAB94698.1"
/db_xref="GI:21204000"
/translation="MNRQNDLILPAVAVIIFPMVWVGFSLFLAGHTPGGPGVGL
LFASLVITTAEDIEIMKRIPIPLDFKILIGLGLVCIATPIASWFLGKNFTHTFD
IPLFILEPVHMTAVFDFGVLCVAVGTMTIISIGENE"
```

Alignment Scores:

Pred. No.:	7.03e+03	Length:	290150
Score:	48.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0

US-10-008-355-26 (1-9) x AP004825 (1-290150)

Qy 1 ThrGlyGlyAsnSerGlySerProval 9

Db 118077 ACTGGTGAACCTACAGTTCACCTGTA 118051

|||||

RESULT 9	AP003132	298050 bp	DNA	linear	BCT 02-JUL-2002
LOCUS	Staphylococcus aureus subsp. aureus N315 genomic DNA, complete				
DEFINITION	genome, section 4/10.				
ACCESSION	AP003132	BA000018			
VERSION	AP003132.2	GI:14349174			
KEYWORDS					
SOURCE					

Staphylococcus aureus subsp. aureus N315 (sub-species:aureus N315, strain:N315) DNA.
Staphylococcus aureus subsp. aureus N315
Bacteria; Firmicutes; Bacillales; Staphylococcus.

REFERENCE 1
AUTHORS Kuroda,M., Ohta,T., Uchiyama,I., Baba,T., Yuzawa,H., Kobayashi,T., Cui,L., Oguchi,A., Aoki,K., Nagai,Y., Lian,J., Ito,T., Kanamori,M., Matsunaru,H., Maruyama,A., Murekami,H., Hosoyama,A., Mizutani-Ui,Y., Takahashi,N.K., Sawano,T., Inoue,R., Kaito,C., Sekimizu,K., Hirakawa,H., Kuhara,S., Goto,S., Yabuzaki,J., Kanehisa,M., Yamashita,A., Oshima,K., Furuya,K., Yoshino,C., Shiba,T., Hattori,M., Ogasawara,N., Hayashi,H. and Hiramoto,K.
Whole genome sequencing of methicillin-resistant Staphylococcus aureus

JOURNAL Lancet 357 (9264), 1225-1240 (2001)

MEDLINE 21311952

PUBMED 11418146

REFERENCE 2 (bases 1 to 298050)

AUTHORS Aoki,K., Oguchi,A., Hosoyama,A., Nagai,Y., Kuroda,M., Hiramatsu,K. and Kikuchi,H.

TITLE Direct Submission
JOURNAL Submitted (30-JAN-2001) Akio Oguchi, National Institute of Technology and Evaluation, Biotechnology Center, ZChome 49-10

Technology and Evaluation, Biotechnology Center, ZChome 49-10
Mishihara, Shihuya-Ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/,
Tel:81-3-3481-8423, Fax:81-3-3481-8424)

COMMENT On Jun 12, 2001 this sequence version replaced gi:13700734.
FEATURES Location/Qualifiers

1..298050
/organism="Staphylococcus aureus subsp. aureus N315"

```
/strain="N315"
/sub_species="aureus N315"
/db_xref="taxon:158879"
280..432
/gene="SA0792"
280..432
```

gene
CDS
/gene="SA0792"
/note="ORFID:SA0792"
/codon_start=1

transl_table=11
product="hypothetical protein"

protein_id="BAB42031.1"
db_xref="GI:13700735"

translation="MKSRSKQPPNRYEAFKPYLLTLTYLAIFITLYLYGSDTHNN
FIYNEF"

448..1905
/gene="q1ta"

448..1905
/gene="q1ta"
/note="ORFID:SA0793"

transl_table=11
product="D-alanine-D-alanyl carrier protein ligase"

protein_id="BAB42032.1"
db_xref="GI:13700736"

translation="MTDIINKLOAFADANPOSIAVRHTTDELTVOQLMDESSKLAHRL
OGSKPMILFGHSPMYMIGMIGATKAGCGYVPDSTIPEDRIKMIINXQPEPENT
TDESRESLEGVFTTEDIKTSODPIEPSQIKNDNTYTTITSSGTGPKVOLEYAS
LVQFTWMLELKSNGKQOMLQAFPSFDLSVMAIYPCLAGGTLNLDKMMINPKL
LNEMLTATPININWSTPSMEMCLPLTNEEOGSLNEFECEGILPHRAKMAIVSR
EPSATIVNTYGPTEATVAVTSIOITQELIDQYPTLPVGVSRGARLSTDDGELVIEG
OSVSLGYLKNQOKTAEVFNDDGIRTYTHGDKARFENGOMFIOGRIDQIKNGYRME
LEETITOLROSEFEVKEAIVPVYKNDKVILHGAIVPTTEVTDNAEMTKNIKNDLSKR
LPEYMIIPKRFEMEQPLPTFSNCKIDRKIAEYING"

1902..3116
/gene="q1tb"

1902..3116
/gene="q1tb"

note="ORFID:SA0794"

transl_table=11
product="D1tb membrane protein"

protein_id="BAB42033.1"
db_xref="GI:13700737"

translation="MIPYGFPEFLIALILPVIILGLKRSYIYGVYAFMIVL
IFSSDKHMLFDQKYISNQLISFIYVWQVILIMFYYSKRKNSFSKPYVMVISTL
ELAYKVQSTWLGCHQIHFRHSKILIEYVFGISYVYFKSVQVLMETRDSIKETIKV
WKLIOFISFPPTISSGPIDRRKRFYKDKKVPVTGNEYRELVAKAHIMLDGLFKYIV
AYFINTVAIMPLQDLHGFVNLMLVWVSYLFFDFAGYSLFAIAFSYLFGLIKTPPN
FDKPFKANINDPFNRMHMTLSFWRDCIYKRSLEFYSRKLLSQFAMSVVAFLNF
FIMGIMHCIEVYIYVGYHAAFLIGYGYERWRRKKHPPRMQNGFTTALSIVTFHFV
TEGFLIFSGKII"

3134..3370
/gene="q1tc"

3134..3370
/gene="q1tc"

note="ORFID:SA0795"

codon_start=1
transl_table=11

product="D-alanyl carrier protein"

protein_id="BAB42034.1"
db_xref="GI:13700738"

translation="MEFREOVUNLLAEVAENDIVKENPDVEIFEGCIDSQVQGLL
EIQKLDIEVSIIMDFDRDMATPKIYALBELR"

3367..4542
/gene="q1td"

3367..4542
/gene="q1td"

note="glycerophosphate chain"

ORFID:SA0796"

codon_start=1
transl_table=11

product="poly D-alanine transfer protein"

/protein_id="BAB42035.1"
/db_xref="GI:13700739"
/translation="MKLKEPFLILSGAVEYFIELLLPASWFTGLVNEKTVEDNRSTLT
DVLKGTIDKGLYESNKKYPIYGSSEIGKDDPPNPALINHNANKKAFILGAGSST
TLINVELASQYDKLKGKLTETIISQWFTNGLTNQNDARNSQTOINOMQOKMS
TLKRYIAORLQFPHVHNKEYLSTAKNPKETKDSYISGFKENOLIKLEAKSLFAM
DKSPLEHVPAKTPDASWDMQAKAVEIKADTYSKRGIRDOYWLILIOESKVRAR
YEFNNSPEFODLELVKTRAAAGADVQVYSIPSNQVWDHIGIDKEROAVYKKIHS
TYVNDGKRIYDMDKDEKYEKVISDAVHIGMKGVVYDEQIAKHMKGEPPQEVDPKRN"
complement(4805..5047)
/gene="SA0797"
complement(4805..5047)
/gene="SA0797"
/note="ORFID:SA0797"
hypothetical protein, similar to nitrogen fixation protein
NifU"
/codon_start=1
/transl_table=11
/protein_id="BAB42036.1"
/db_xref="GI:13700740"
/translation="MPTEDTMDQVAEYIERLRPEPLLDGSGDCLIDVEDGIYKQL
HGACGSPSSSTTLKAGIERALHEEPVIEYQVF"
5163..5471
/gene="SA0798"
5163..5471
/gene="SA0798"
/note="ORFID:SA0798"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42037.1"
/db_xref="GI:13700741"
/translation="MYYGADYICASCVNAPNSKDIYDLQPLKRYKNISFKYTYI
DITKNDNTIDHDQFIERIDDELFLPITANDEVADGYIOTKOTITFIDOKLNE
"
complement(5530..6594)
/gene="SA0799"
complement(5530..6594)
/gene="SA0799"
/note="ORFID:SA0799"
hypothetical protein, similar to MDH dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB42038.1"
/db_xref="GI:13700742"
/translation="MKNIVLLGGYGNMRINSRLTSLPQDYTVTVLRMPFHGKIP
EYVALAGTSKDVDRMKNPNOVNTVGEINDIDLDKQIVSFGNSKIDYDELLIGL
GGEQRYHNYPGAEETHSIQTLISKARDTFHSISEPEGAKVIGAGLSGIELASERL
ESRSDLEIYLRGPRILRNPEKLSKAYKFNKANNVAVNSNINKPEPGYICD
BKRIDLVYWTAGIQPVEVVRNLPIDINSNGRVINQHVPTVRNVYVGCADLPH
APSAQLAEVQGGDIADVLKQWLNPLPKMPELAKVQIGVSLGDKGFAIYIMDRIVT
GLASLTKSGVLMIKYHNG"
6911..7147
/gene="SA0800"
6911..7147
/gene="SA0800"
/note="ORFID:SA0800"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42039.1"
/db_xref="GI:13700743"
/translation="MNPIVEFLTSNMAGKGDVFNQLENDDPVVDLEYGCLTHGIGCS
AGLYALNGDIEGDSPELLQNIYAHIKETWIF"
7160..7519
/gene="SA0801"
7160..7519
/gene="SA0801"
/note="ORFID:SA0801"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAB42040.1"

/db_xref="GI:13700744"
/translation="MPYVILTEAAAYEVKMLKANMPPGYLIKIKVNGGCTGLTYGM
GAEAPAGNDEVILEYFGLKVLVDKEDAPVNLNGTTIDFKSLMGGGQIDNPANISCG
CCSSPFTAVAGNPENC"
7973..9181
/gene="SA0802"
7973..9181
/gene="SA0802"
/note="ORFID:SA0802"
hypothetical protein, similar to MDH dehydrogenase"
/codon_start=1
/transl_table=11
/protein_id="BAB42041.1"
/db_xref="GI:13700745"
/translation="MAQDRKYIVAGYAGIQTYYKLOKATSTEAETILINKNEYH
YEATWLHESAGTLNEDLYEVESYKDKKNFVOAETKIDRDKKFTNQTGTYDE
DILVVALGVSEFTGIEGMDHAFQIENVITARELSRHLDEKFAVNAASKEKDDNLS
ILVGAFTGVFELGELTDRIPELDSKYGVDPQNKVYITCVEAPKWLPMFSEELVSHA

Alignment Scores:
Pred. No.: 7,16e+03 Length: 298050
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-26 (1-9) x AP003132 (1-298050)
OY 1 ThrGlyGlyAsnSerGlySerProval 9
Db 123979 ACTGGTGTAACTACAGCTTCACCTGTA 123953
|||||
RESULT 10
AP003361/c 347235 bp DNA linear BCT 07-FEB-2002
LOCUS
DEFINITION
Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
sequence, section 4/9.
ACCESSION
AP003361 BA000017
VERSION
AP003361.2 GI:114246761
KEYWORDS
SOURCE
Staphylococcus aureus subsp. aureus Mu50 (sub-species:aureus Mu50,
strain:Mu50) DNA.
ORGANISM
Staphylococcus aureus subsp. aureus Mu50
Bacteria: Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1 Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I.,
Cui, L., Oguchi, A., Aoki, K., Nagai, Y., Iino, J., Ito, T., Kanamori, M.,
Matsumaru, H., Maruyama, A., Murakami, H., Hosoyama, A.,
Mizutani, U., Y., Takahashi, N. K., Sawano, T., Inoue, R., Kaito, C.,
Sekimizu, K., Hirakawa, H., Kohara, S., Goto, S., Yabuzaki, J.,
Kanehisa, M., Yamashita, A., Oshima, K., Furuya, K., Yoshino, C.,
Shiba, T., Hattori, M., Ogasawara, N., Hayashi, H. and Hiramatsu, K.
Whole genome sequencing of methicillin-resistant Staphylococcus
aureus
JOURNAL Lancet 357 (9264), 1225-1240 (2001)
MEDLINE 21311952
REFERENCE 2 (bases 1 to 347235)
AUTHORS Ohta, T.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-2001) Toshiko Ohta, University of Tsukuba College
of Medical Technology and Nursing, Department of Medical
Technology, 1-1-1 Ten-nodai, Tsukuba, Ibaraki 305-8577, Japan
(E-mail: tohhasakura.ac.tsukuba.ac.jp, Tel:81-298-53-3454,
Fax:81-298-53-3454)
COMMENT On May 29, 2001 this sequence version replaced gi:113875305.
FEATURES
1..347235
location/Qualifiers
source
/organism="Staphylococcus aureus subsp. aureus Mu50"
/strain="Mu50"
/sub_species="aureus Mu50"
/db_xref="taxon:158878"
35..1021
gene

```

CDS
    /gene="SAV0992"
    35..1021
    /gene="SAV0992"
    /codon_start=1
    /transl_table=11
    /product="oligopeptide ABC transporter ATP-binding protein
    homologue"
    /protein_id="BAB57154.1"
    /db_xref="GI:14246762"
    /translation="MNNVLEFVKDLETSTIKINEMVLATVENISFELSGEVLGTVGES
    GCGKSLKSTIKLPEKTSKSGEVITDGRKIDITLNKQULDIRGNDIATIFDEPM
    TALNPFYTIKQLVESIKSHKISKREANNLAKDLKKGIAKODEILNSYPHOLSG
    MKQRYWIAWISCPRLIADDEPTALDVTIQAILDLKLEKQETOMAINMITLSD
    VVAEFCQKLVYAGQIVGFGIKELIHNPKHPYQKLSITPKLEKQKRLTEIEGI
    VPSIAFHVNCRFANRCNKKLDICNNSPKMHVCEQDIVRCHLYKNEKKE"
    1024..2004
    /gene="appf"
    /gene="appf"
    1024..2004
    /gene="SAV0993"
    /note="SAV0993"
    /codon_start=1
    /transl_table=11
    /product="oligopeptide transport system ATP-binding
    protein Appf homologue"
    /protein_id="BAB57155.1"
    /db_xref="GI:14246763"
    /translation="MENILEVNOIKKYYIKTGLQKQYKAVDVSFIKKGQTFG
    LVGSGCGKSTTGKVIIRLEDAISGSIYNGEDITRLOCKIKRKSROOQMTFQPIYA
    SINPMQWVDITSEPIINRKLPKEIKKEVYLYLKVGLSEDAITKIAHESGQRO
    KGIANALALRPSLIADDEPVASLDVSOVNLNLKDLQEQFKLSYFIADLSYVK
    HSDVIGVYLGHIVEIASDKIYEMPKHPYKALISSIPQIDKHNNNNIILKGEIPLS
    PSNPQGCFFHRCPIAKDKCKENIPQLKDIGNEHQVACFYVDKVDLNA"
    1997..2959
    /gene="oppb"
    /gene="oppb"
    1997..2959
    /gene="SAV0994"
    /note="SAV0994"
    /codon_start=1
    /transl_table=11
    /product="probable oligopeptide transport system permease
    protein"
    /protein_id="BAB57156.1"
    /db_xref="GI:14246764"
    /translation="MKLIKRLGIMTPLLILISIVESLAIQPGDPFSDLQNKIK
    QEAINAQREKLNDISIPHOYIRWNVHVGDLGESIKKRPVIDIEIRIPNTILG
    ASLITITISFALGITSGRYSILDVQVIFNLMLAIPESIAVGAFIIFESFELD
    WPPQGSVDINKKEGFEFYMSKIYHTFIPATIGLSAGIQTURLNDIENSKDY
    VITANSKLSMKRTINKHILNLSLPIITFFLGADIVSILGSAVITETIFSYNGICKLF
    LESVIGQDPLMKALTLEFFSLGLIGNLISDIYGFIDPRIRSN"
    2971..3852
    /gene="SAV0995"
    /gene="SAV0995"
    2971..3852
    /gene="SAV0995"
    /note="similar to oligopeptide transport system permease
    protein Oppc"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAB57157.1"
    /db_xref="GI:14246765"
    /translation="MOKSKSPFKIALSFHINKIAMLSIIFLLITITVSIAPLIAP
    FVNOODLNINGEMTAONILGDSGRNFSRLYAGRISLSIGITSTIGMLLIGIT
    VVSIQSGFQIVDTLMTTEFVMLPEPLFALIVNALGDKIKNPYGAIIILVVIIT
    VLSMGILARVKGVLQEKENEYPLAAKSGTPTKTIILKHLLPNLISVYIQAILLF
    AGMIYVESLSLFGFISKAIRPSWGMNLSDAQEGVISGKPMIMPALMITLITLISI
    NTVGGLKDAFNPGRK"
    complement(3894..4883)
    /gene="trps"
    complement(3894..4883)
    /gene="trps"
    /note="SAV0996"
    /codon_start=1
    /transl_table=11
    /gene="SAV0997"
    /gene="SAV0997"
    5177..5572
    /gene="SAV0997"
    5177..5572
    /gene="SAV0997"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="BAB57159.1"
    /db_xref="GI:14246767"
    /translation="WVLTFSPTSCYSCRKAKMLQEHDIPTERNIFSEHLITDEIKQ
    LKMTEDGDEIISTRSKYORKLVNDISLPLQDLYSIIQDPGLRRPITLDNKRLO
    VCYNDEILRRPLPRKVRTPFQLEAQRMD"
    5943..6662
    /gene="SAV0998"
    /gene="SAV0998"
    5943..6662
    /gene="SAV0998"
    /note="similar to negative regulator of genetic competence
    Meca"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAB57160.1"
    /db_xref="GI:14246768"
    /translation="MRIRVDQDTYKLFITYSIDIEARSGFREDLTNRKRGPEFWSM
    MDEINBEDFYVEGPMQVYAFKEGVETITSKSNEDMANNSDDATDQDEQVQL
    LAQILEGQULEBELQRTKEKAQSGRQSSAKNRTIIVKNDLEDVINTAVHS
    NPITTEFDLLYMDGTYYAVAFPSHVDQEVINDISYQLLEFAVPTORTVEYLNDYA
    KIMSHNVTAQVRRYPETTE"
    6783..7769
    /gene="SAV0999"
    /gene="SAV0999"
    6783..7769
    /gene="SAV0999"
    /note="similar to transcription factor"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAB57161.1"
    /db_xref="GI:14246769"
    /translation="MLVALNEKERVLAFTTALRKTQYCPVCGKQVILKRLGVISHF
    AHKHLAEKCFENNETIKHYKSKLILAOIMIQOQGVLEPEFLKEIKQIPDILINNKYV
    IELQYSPIPYKQIILQRTGLKMGYKVSMLNDVDYCHNKKVFNHSHLFTNPTRKL
    HTEFLKKOIMHFOOIOYLGCHKYVAEKRNAKIIELFMDACDQYHAYVKKLSPATNOY
    IKYCRWQNSVLEPTISAMYQDLTQDEVYVNGYTFPPQIYIENRPIEMQLOVDMLK
    NKSRLVNDNLNTEFLKRFVLALESKTAIIELKLNINYNICSDRNDVQILF"
    7817..9625
    /gene="SAV1000"
    /gene="SAV1000"
    7817..9625
    /gene="SAV1000"
    /codon_start=1
    /transl_table=11
    /product="thimet oligopeptidase homologue"
    /protein_id="BAB57162.1"
    /db_xref="GI:14246770"
    /translation="MSQSLRSREOGRKPYETWDLTITFKNDEAFAPKFEYENELGK
    EPOFGHIGDSKETIYNLLEEDTIGTLEKYYVAAHKQODDNTNDYTGESARQ
    LIKFSAMSELPPELIDIDEDKIOSFVNSIDKDLQKFAFDLKLNEKRPHTLIDATEK
    LITEAQDALSTPSNYYGMSNADLFEADIDKGAHPLTQGTPIKLYLESDDRKLRES
    AFKNYKAYGANNNTLGAIVAGEVKNFANRTNAYKTAREKALSNNHPIENVYDNLV
    KTVHKYLLPLHRYTELKREKLEGLDGLKAYDLYTPIIKIDKEFMEPEAKEMWLKLEP
    MGEELYLVVKGGLNNRWVDYVENKGRSGGYSAGHLLNPILNMSNTISDLYLVH
    EPGSHAHSEFSKRPQPSMSDYTITVAAYASPCNALSISVDNRKHLDDERKLLINQEL
    LERFRATLFRQTFMFAEFHKIHAIEAGEEPLTPTRMNEEIVAKLNLIFGDSVETEDI
    SKEMSRIPHFMNYYVQYATGYSAQSLSHQILTEGKPAVDRIYNEPLKKGSSNYPPI
    EILKNAGVDMTTPPEIQACEVFQGLNAFEKLMKA"

```

```

gene
    /product="tryptophanyl-tRNA synthetase"
    /protein_id="BAB57158.1"
    /db_xref="GI:14246766"
    /translation="METLFSGIQPSGCIPTTIGNYICALKQFPVQVNDYDCIYCIQDHA
    ITMPDRLKLRQTLQALAIYIASIDDPKATLFIQSVSPAHVQGMMLTTLIASGEL
    ERMTOYKDKAKQAVEGIPAGLITTPPLMAQVLTNTNIVPQDQKQKHIELTRNVLVD
    RENSRYNDVLVLEIDPEIRMPKSGVGRKSNKAKNFISLDEPVAKAKI
    KSAVDSQDGIIEKFDNRNKGITNLISITAGLPLDMFTIKDIEAKYEEGEGKFGKDLAEI
    VKAFLVEQERKESEFYSNDDKLDLIDQGRDAHAKVSPFYVKMKEMAKGLCKRR"
    5177..5572
    /gene="SAV0997"
    /gene="SAV0997"
    5177..5572
    /gene="SAV0997"
    /codon_start=1
    /transl_table=11
    /product="conserved hypothetical protein"
    /protein_id="BAB57159.1"
    /db_xref="GI:14246767"
    /translation="WVLTFSPTSCYSCRKAKMLQEHDIPTERNIFSEHLITDEIKQ
    LKMTEDGDEIISTRSKYORKLVNDISLPLQDLYSIIQDPGLRRPITLDNKRLO
    VCYNDEILRRPLPRKVRTPFQLEAQRMD"
    5943..6662
    /gene="SAV0998"
    /gene="SAV0998"
    5943..6662
    /gene="SAV0998"
    /note="similar to negative regulator of genetic competence
    Meca"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAB57160.1"
    /db_xref="GI:14246768"
    /translation="MRIRVDQDTYKLFITYSIDIEARSGFREDLTNRKRGPEFWSM
    MDEINBEDFYVEGPMQVYAFKEGVETITSKSNEDMANNSDDATDQDEQVQL
    LAQILEGQULEBELQRTKEKAQSGRQSSAKNRTIIVKNDLEDVINTAVHS
    NPITTEFDLLYMDGTYYAVAFPSHVDQEVINDISYQLLEFAVPTORTVEYLNDYA
    KIMSHNVTAQVRRYPETTE"
    6783..7769
    /gene="SAV0999"
    /gene="SAV0999"
    6783..7769
    /gene="SAV0999"
    /note="similar to transcription factor"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAB57161.1"
    /db_xref="GI:14246769"
    /translation="MLVALNEKERVLAFTTALRKTQYCPVCGKQVILKRLGVISHF
    AHKHLAEKCFENNETIKHYKSKLILAOIMIQOQGVLEPEFLKEIKQIPDILINNKYV
    IELQYSPIPYKQIILQRTGLKMGYKVSMLNDVDYCHNKKVFNHSHLFTNPTRKL
    HTEFLKKOIMHFOOIOYLGCHKYVAEKRNAKIIELFMDACDQYHAYVKKLSPATNOY
    IKYCRWQNSVLEPTISAMYQDLTQDEVYVNGYTFPPQIYIENRPIEMQLOVDMLK
    NKSRLVNDNLNTEFLKRFVLALESKTAIIELKLNINYNICSDRNDVQILF"
    7817..9625
    /gene="SAV1000"
    /gene="SAV1000"
    7817..9625
    /gene="SAV1000"
    /codon_start=1
    /transl_table=11
    /product="thimet oligopeptidase homologue"
    /protein_id="BAB57162.1"
    /db_xref="GI:14246770"
    /translation="MSQSLRSREOGRKPYETWDLTITFKNDEAFAPKFEYENELGK
    EPOFGHIGDSKETIYNLLEEDTIGTLEKYYVAAHKQODDNTNDYTGESARQ
    LIKFSAMSELPPELIDIDEDKIOSFVNSIDKDLQKFAFDLKLNEKRPHTLIDATEK
    LITEAQDALSTPSNYYGMSNADLFEADIDKGAHPLTQGTPIKLYLESDDRKLRES
    AFKNYKAYGANNNTLGAIVAGEVKNFANRTNAYKTAREKALSNNHPIENVYDNLV
    KTVHKYLLPLHRYTELKREKLEGLDGLKAYDLYTPIIKIDKEFMEPEAKEMWLKLEP
    MGEELYLVVKGGLNNRWVDYVENKGRSGGYSAGHLLNPILNMSNTISDLYLVH
    EPGSHAHSEFSKRPQPSMSDYTITVAAYASPCNALSISVDNRKHLDDERKLLINQEL
    LERFRATLFRQTFMFAEFHKIHAIEAGEEPLTPTRMNEEIVAKLNLIFGDSVETEDI
    SKEMSRIPHFMNYYVQYATGYSAQSLSHQILTEGKPAVDRIYNEPLKKGSSNYPPI
    EILKNAGVDMTTPPEIQACEVFQGLNAFEKLMKA"

```

gene complement(10085, .10639)
/gene="y1bm"
CDS complement(10085, .10639)
/gene="y1bm"
/note="SAV1001"
/codon_start=1
/transl_table=11
/product="GTP pyrophosphokinase"
/protein_id="BAB57163.1"
/db_xref="GI:14246771"
/translation="MQSLKSYVSKTEKTYTCTNEIIPKRDITESHICDIONAG
IDLEVFKDQKSKLSESLKIDLHAREMEIQADSLVFESEVHEGLKVGGLPYH
ITYEILINELMGKPIEKNLPKLEIYIQOQOLVTMELITLYEMPEKILNELKLAIQ
OKIERLKYPDGDPMKSKMKIRSK"
complement(10575, .10901)
/gene="SAV1002"

Alignment Scores:
Pred. No.: 7.93e+03 Length: 347235
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-26 (1-9) x AF003361 (1-347235)

QY 1 ThrGlyGlyAsnSergGlySerProVal 9
|||||
Db 56369 ACTGGTGGTAACTGAGTTCACCTGTA 56343

RESULT 11
AF008576/c 1912 bp DNA linear PRI 05-FEB-1998
LOCUS Homo sapiens galanin receptor gene, 5'UTR.
DEFINITION AF008576
ACCESSION AF008576.1 GI:2352810
VERSION
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1912)
AUTHORS Lorimer,D.D., Matkowski,K. and Benya,R.V.
TITLE Cloning, chromosomal location, and transcriptional regulation of
the human galanin-1 receptor gene (GAN1R)
JOURNAL Biochem. Biophys. Res. Commun. 241 (2), 558-564 (1997)
MEDLINE 98086390
PUBMED 9425310
REFERENCE 2 (bases 1 to 1912)
AUTHORS Lorimer,D.D., Lui,W., Matkowski,K. and Benya,R.V.
TITLE Direct Submission
JOURNAL Submitted (16-JUN-1997) Medicine, University of Illinois at
Chicago, 840 S. Wood St. (M/C 787), Chicago, IL 60612, USA
FEATURES
source 1..1912
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="13q24"
1..1912
/note="galanin receptor gene"

BASE COUNT 460 a 498 c 573 g 381 t
ORIGIN

Alignment Scores:
Pred. No.: 950 Length: 1912
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-26 (1-9) x AF008576 (1-1912)

QY 1 ThrGlyGlyAsnSergGlySerPro 8
|||||
Db 1210 ACTGGTGGTAACTGAGTTCACCTGTA 1187

RESULT 12
AC108022 162609 bp DNA linear PRI 20-MAR-2002
LOCUS Homo sapiens BAC clone RP11-9D8 from 4, complete sequence.
DEFINITION AC108022 AC009566
ACCESSION AC108022.3 GI:19339101
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 162609)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 162609)
AUTHORS Cedroni,M., Kozlowski,A. and Spalding,L.
TITLE The sequence of Homo sapiens BAC clone RP11-9D8
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 162609)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 162609)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 162609)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 162609)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Mar 11, 2002 this sequence version replaced gi:18693535.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0009D08
Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tateo, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-36B15; the clone sequenced to the right is RP11-177C12. 2000 bp overlap. Actual end of this clone is at base position 12410 of RP11-177C12.

Polymorphisms have been identified between AC108022 and AC021106.

FEATURES
The sequence of AC009566 has been incorporated into AC108022.

source
Location/Qualifiers

1. 162609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-9D8"
/clone.lib="RPCT-11"
40. 353
/rpt_family="Alu"
/rpt_family="Alu"
329. 510
/note="match to EST AW816997 (NID:97909991)"
344. 521
/note="match to EST BE158378 (NID:98621099)"
828. 1376
/note="match to EST BF898400 (NID:912289859)"
1063. 1248
/note="match to EST BF734110 (NID:912059346)"
1299. 1719
/note="similar to Homo sapiens EST BE069322 (NID:98413972)"
1495. 1929
/note="match to EST A1828853 (NID:95449524) ts81905.x1"
2267. 2350
/rpt_family="MIR"
2389. 2749
/note="similar to Homo sapiens EST B1020745 (NID:914423753)"
2689. 2836
/rpt_family="MER1_type"
3066. 3203
/rpt_family="L2"
3705. 3816
/rpt_family="L2"
3859. 3952
/rpt_family="MIR"
4170. 4271
/rpt_family="MIR"
4513. 4550
/rpt_family="MALR"
4551. 4849
/rpt_family="Alu"
4850. 5239
/rpt_family="MALR"
5554. 5682
/rpt_family="MIR"
6188. 6415
/note="match to EST BE172154 (NID:98634880)"

repeat_region 6310. 6342
/rpt_family="L1"
repeat_region 6343. 6534
/rpt_family="MER2_type"
repeat_region 6535. 6830
/rpt_family="Alu"
repeat_region 6831. 6948
/rpt_family="MER2_type"
repeat_region 6949. 7453
/rpt_family="L1"
repeat_region 7522. 7815
/rpt_family="Alu"
repeat_region 8292. 8347
/rpt_family="MIR"
repeat_region 8381. 8688
/rpt_family="Alu"
repeat_region 8827. 9065
/rpt_family="MIR"
repeat_region 10185. 10227
/rpt_family="ERV1"
repeat_region 10225. 10336
/rpt_family="MIR"
repeat_region 10337. 10648
/rpt_family="Alu"
repeat_region 10649. 10661
/rpt_family="MIR"
repeat_region 10662. 10966
/rpt_family="Alu"
repeat_region 10967. 11016
/rpt_family="MIR"
repeat_region 11533. 11838
/rpt_family="Alu"
misc_feature 12642. 13853
/note="CPG_Island (%GC=73.3, o/e=0.90, #CPGs=137)"
misc_feature 13062. 13216
/note="match to EST BE792154 (NID:910213352)"
repeat_region 13268. 13339
/rpt_family="CCCCG)n"
repeat_region 14275. 14373
/rpt_family="L2"
repeat_region 14952. 15119
/rpt_family="ACHobo"
repeat_region 15257. 15567
/rpt_family="Alu"
repeat_region 15645. 15745
/rpt_family="MIR"
repeat_region 15849. 15950
/rpt_family="GA-rich"
repeat_region 16260. 16429
/rpt_family="MIR"
repeat_region 17155. 17455
/rpt_family="Alu"
repeat_region 17462. 17607
/rpt_family="MIR"
repeat_region 17656. 17864
/rpt_family="MER1_type"
repeat_region 17867. 17961
/rpt_family="L2"
repeat_region 18268. 18395
/rpt_family="Alu"
repeat_region 19184. 19456
/rpt_family="Alu"

Alignment Scores:

Pred. No.: 1.86e+04 Length: 162609
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-26 (1-9) x AC108022 (1-162609)

QY 1 ThrglyGlyansserglyserPro 8
|||||
Db 42377 ACTGTCGAAATCTCGATCCCA 42400

RESULT 13
AC108085/c
LOCUS
DEFINITION
AC108085 166978 bp DNA linear HTG 25-JAN-2002
Homo sapiens chromosome 5 clone CTD-2210F7, WORKING DRAFT SEQUENCE,
6 unordered pieces.
AC108085
VERSION
AC108085.1 GI:18369931
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 166978)
DOE Joint Genome Institute.
Unpublished
2 (bases 1 to 166978)
DOE Joint Genome Institute.
Direct Submission
Submitted (25-JAN-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 708316
Center clone name: CITB-H1_2210F7

Summary Statistics
Consensus quality: 163223 bases at least Q40
Consensus quality: 164704 bases at least Q30
Consensus quality: 165029 bases at least Q20
Estimated insert size: 179000; agarose-fp estimation
Estimated insert size: 166478; sum-of-contigs estimation
Quality coverage: 8.17 in Q20 bases; agarose-fp estimation
Quality coverage: 8.78 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1111: contig of 1111 bp in length
* 1112 1211: gap of unknown length
* 1212 2531: contig of 1320 bp in length
* 2532 2632: gap of unknown length
* 2632 4084: contig of 1453 bp in length
* 4085 4185: gap of unknown length
* 4185 50656: contig of 46472 bp in length
* 50657 50756: gap of unknown length
* 50757 101216: contig of 50460 bp in length
* 101217 101316: gap of unknown length
* 101317 166978: contig of 65662 bp in length.
Location/Qualifiers
1. 166978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone_lib="CTD-2210F7"
/clone_lib="Caltech human BAC library D"
ORIGIN
53321 a 32702 c 32339 g 48053 t 563 others
Alignment Scores:
Pred. No.: 1.89e+04 Length: 166978

Score: 44.00
Percent Similarity: 88.89%
Best Local Similarity: 88.89%
Query Match: 91.67%
DB: 2
Gaps: 0

US-10-008-355-26 (1-9) x AC108085 (1-166978)

QY 1 ThrglyGlyansserglyserProval 9
|||||
Db 2529 ACAAGGGGCAATCTCGGCGCCGTT 2503

RESULT 14
AC036185
LOCUS
DEFINITION
AC036185 171279 bp DNA linear HTG 21-JUN-2000
Homo sapiens chromosome 4 clone RP11-67986 map 4, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
AC036185
VERSION
AC036185.2 GI:8576273
HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 171279)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 4, clone RP11-67986
Unpublished
2 (bases 1 to 171279)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
Campiano,A., Castle,A., Choquel,X., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hages,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Landers,T., Lebecky,J.,
Levine,R., Liu,G., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeckers,R.,
Meldrum,J., Menes,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 21, 2000 this sequence version replaced g1:7523854.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9247
Center clone name: 679_B_6
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 163254 bases at least Q40
Consensus quality: 166954 bases at least Q30
Consensus quality: 168680 bases at least Q20
Insert size: 170000; agarose-fp

Insert size: 169879; sum-of-contigs
Quality coverage: 7.1 in Q20 bases; agarose-fp
Quality coverage: 7.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1385: contig of 1385 bp in length
* 1386 1485: gap of 100 bp
* 1486 2835: contig of 1350 bp in length
* 2836 2935: gap of 100 bp
* 2936 4108: contig of 1173 bp in length
* 4109 4208: gap of 100 bp
* 4209 7507: contig of 3299 bp in length
* 7508 7607: gap of 100 bp
* 7608 13869: contig of 6262 bp in length
* 13870 13969: gap of 100 bp
* 13970 20810: contig of 6841 bp in length
* 20811 20910: gap of 100 bp
* 20911 28893: contig of 7983 bp in length
* 28894 28993: gap of 100 bp
* 28994 39086: contig of 10093 bp in length
* 39087 39186: gap of 100 bp
* 39187 50077: contig of 10891 bp in length
* 50078 50177: gap of 100 bp
* 50178 61526: contig of 11349 bp in length
* 61527 61626: gap of 100 bp
* 61627 75529: contig of 13903 bp in length
* 75530 75629: gap of 100 bp
* 75630 91544: contig of 15915 bp in length
* 91545 91644: gap of 100 bp
* 91645 108750: contig of 17106 bp in length
* 108751 108850: gap of 100 bp
* 108851 136877: contig of 28027 bp in length
* 136878 136977: gap of 100 bp
* 136978 171279: contig of 34302 bp in length.

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-67986"
/clone_lib="RP11 Human Male BAC"
1.1385
/note="assembly_fragment"
misc_feature
1486.2835
/note="assembly_fragment"
misc_feature
2936.4108
/note="assembly_fragment"
misc_feature
4209.7507
/note="assembly_fragment"
misc_feature
7608.13869
/note="assembly_fragment"
misc_feature
13970.20810
/note="assembly_fragment"
misc_feature
20911.28893
/note="assembly_fragment"
misc_feature
28994.39086
/note="assembly_fragment"
misc_feature
39187.50077
/note="assembly_fragment"
misc_feature
50178.61526
/note="assembly_fragment"
misc_feature
61627.75529
/note="assembly_fragment"
misc_feature
136978.171279
vector_side:right"

misc_feature 75630..91544
/note="assembly_fragment"
misc_feature 91645..108750
/note="assembly_fragment"
misc_feature 108851..136877
/note="assembly_fragment"
misc_feature 136978..171279
/note="assembly_fragment"
BASE COUNT 48027 a 36049 c 36303 g 49499 t 1401 others
ORIGIN

Alignment Scores:

Pred. No.: 1.93e+04 Length: 171279
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 2 Gaps: 0

US-10-008-355-26 (1-9) x AC036185 (1-171279)

QY 1 ThrGlyGlyAsnSerGlySerPro 8

Db 126420 ACTGTGGAATTCGTGATCCCA 126443

RESULT 15

AC068590

LOCUS 174253 bp DNA linear HTG 24-AUG-2002
DEFINITION Homo sapiens chromosome 8 clone RP11-131K16 map 8, WORKING DRAFT

SEQUENCE 30 unordered pieces.

AC068590

AC068590.2 GI:8389591

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

human.
Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 174253)

REFERENCE Birren,B., Linton,L., Nusbaum,C. and Lander,E.

JOURNAL Homo sapiens chromosome 8, clone RP11-131K16

REFERENCE 2 (bases 1 to 174253)

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavsky,L., Boukngalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collymore,A., Cooke,P., Deatellano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,

Galagan,J., Gardyna,S., Glnde,S., Goyette,M., Graham,L.,

Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,

Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,

Klein,J., Laroque,K., Lamazares,R., Landers,T., Lehoczy,J.,

Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,

McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McHeeters,R.,

Meldrim,J., Meheus,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J.,

Murphy,T., Naylor,T., Norman,C.H., O'Connor,T., O'Donnell,P.,

O'Neil,D., Olivari,T.M., Oliver,J., Peterson,K., Pierre,N.,

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,

Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Testave,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,

Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,

Young,G., Zalnoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavsky,L., Boukngalter,B., Brown,A., Burkett,G.,

Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,


```

misc_feature      /note="assembly_fragment"
                  28293..32574
misc_feature      /note="assembly_fragment"
                  32675..36983
misc_feature      /note="assembly_fragment"
                  37084..41557
misc_feature      /note="assembly_fragment"
                  41658..46775
misc_feature      /note="assembly_fragment"
                  46876..52390

```

```

Alignment Scores:
Pred. No.:      1.95e+04      Length:      174253
Score:          44.00         Matches:      8
Percent Similarity: 100.00%   Conservative: 0
Best Local Similarity: 100.00% Mismatches:      0
Query Match:      91.67%      Indels:         0
DB:              2           Gaps:            0

```

US-10-008-355-26 (1-9) x AC068590 (1-174253)

```

Oy      1  ThrGlyGlyAsnSerGlySerPro 8
         |||||
Db 40541 ACAGGAGGAGACAGTGCAGCCCT 40564

```

Search completed: December 20, 2002, 15:04:09
 Job time : 1506.26 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 07:26:30 ; Search time 53.6842 Seconds
(without alignments)
38.381 Million cell updates/sec

Title: US-10-008-355-25

Perfect score: 54

Sequence: 1 TGNNGSPVF 10

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*
15: SP-virus:*
16: SP-bacteriap:*
17: SP-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54	100.0	316	2 Q9FBG1	Q9FBG1 staphylococ
2	54	100.0	342	16 Q99V45	Q99V45 staphylococ
3	54	100.0	357	2 Q04186	Q04186 staphylococ
4	49	90.7	217	2 Q9AJX0	Q9AJX0 staphylococ
5	48	88.9	716	16 Q9PC94	Q9PC94 xyloella fas
6	44	81.5	284	2 Q47809	Q47809 enterococcu
7	44	81.5	289	16 Q98M08	Q98M08 rhizobium l
8	44	81.5	398	11 Q8VE75	Q8VE75 mus musculu
9	44	81.5	613	11 Q9D2L9	Q9D2L9 mus musculu
10	43	79.6	1008	16 Q98K09	Q98K09 rhizobium l
11	42	77.8	169	16 Q8YM87	Q8YM87 anabaena sp
12	42	77.8	458	16 Q8XHM4	Q8XHM4 clostridium
13	40	74.1	315	16 Q9ZDX8	Q9ZDX8 rickettsia
14	40	74.1	353	11 Q62313	Q62313 mus musculu
15	40	74.1	363	11 Q62314	Q62314 mus musculu
16	40	74.1	497	16 Q92J35	Q92J35 rickettsia

17	40	74.1	639	16 Q92LB9	Q92LB9 rhizobium m
18	39	72.2	319	16 Q9A9N9	Q9A9N9 caulobacter
19	39	72.2	364	16 Q8YFB1	Q8YFB1 brucella me
20	39	72.2	379	3 Q9Y891	Q9Y891 collettotric
21	39	72.2	396	12 Q8VA04	Q8VA04 apple stem
22	39	72.2	499	10 Q9LK70	Q9LK70 arabidopsis
23	39	72.2	996	3 Q9P7S1	Q9P7S1 schizosacch
24	39	72.2	997	3 Q74325	Q74325 schizosacch
25	38	70.4	253	2 P96151	P96151 vibrio chol
26	38	70.4	359	16 Q98B89	Q98B89 rhizobium l
27	38	70.4	394	12 Q8QR15	Q8QR15 apple stem
28	38	70.4	440	17 Q8TZS8	Q8TZS8 pyrococcus
29	38	70.4	441	10 Q9FMX1	Q9FMX1 arabidopsis
30	38	70.4	470	16 Q9RTZ8	Q9RTZ8 deinococcus
31	38	70.4	476	16 Q8KLD4	Q8KLD4 vibrio chol
32	38	70.4	568	12 Q89504	Q89504 cocksfoot m
33	38	70.4	568	12 Q66149	Q66149 cocksfoot m
34	38	70.4	568	12 Q9E960	Q9E960 cocksfoot m
35	38	70.4	579	16 Q8RE47	Q8RE47 fusobacteri
36	38	70.4	593	10 Q9FY14	Q9FY14 arabidopsis
37	38	70.4	627	10 Q9SRP2	Q9SRP2 arabidopsis
38	38	70.4	1097	10 Q8RY22	Q8RY22 arabidopsis
39	38	70.4	1335	2 Q9LA58	Q9LA58 escherichia
40	38	70.4	1335	2 Q9LA54	Q9LA54 escherichia
41	37	68.5	235	2 Q9KH51	Q9KH51 staphylococ
42	37	68.5	235	2 Q9FD08	Q9FD08 staphylococ
43	37	68.5	235	16 Q99T60	Q99T60 staphylococ
44	37	68.5	239	2 Q9KH49	Q9KH49 staphylococ
45	37	68.5	239	16 Q53782	Q53782 staphylococ

ALIGNMENTS

RESULT 1
Q9FBG1 PRELIMINARY; PRT; 316 AA.
ID Q9FBG1;
AC Q9FBG1;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Glutamyl endopeptidase.
GN PROM.
OS Staphylococcus warneri.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1292;
RN [1]
RP SEQUENCE FROM N.A.
RA Kakiwara M.;
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kodaira K.I.;
RT "Characterization of the gene encoding glutamyl endopeptidase of
RT Staphylococcus warneri M.";
RL Submitted (Sep-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ293885; CAC06168.1; -
DR MEROPS; S01.269; -
DR InterPro; IPR001254; Ser_protease_Try.
DR InterPro; IPR00126; Ser_proteas_V8.
DR Pfam; PF00089; Trypsin; I.
DR PRINTS; PR00039; V8PROTEASE.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolyase; Serine protease.
SO SEQUENCE 316 AA; 34296 MW; 4E997A5A111DBD40 CRC64;
Query Match 100.0%; Score 54; DB 2; Length 316;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10
|||||
Db 231 TCGNSGSPVF 240

RESULT 2

099V45 PRELIMINARY: PRT: 342 AA.
AC 099V45;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Serine protease, V8 protease, glutamyl endopeptidase.
GN SSPA OR SAV1048 OR SA0901.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus
ON NCBI_TaxID=158878, 158879;
RN [1]

RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Raito C.,
Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
CC -1 SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AP003361; BAB57210.1; -;
DR EMBL: AP003132; BAB42146.1; -;
DR MEROPS: S01.269; -;
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000126; Ser_proteas_V8.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00020; Tryp-Spec; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00672; V8_HIS; 1.
DR PROSITE: PS00673; V8_SER; 1.
KW Hydrolyase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 342 AA; 36977 MW; 5AEF42DCE01C4B24 CRC64;

Query Match 100.0%; Score 54; DB 16; Length 342;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10
|||||
Db 233 TCGNSGSPVF 242

RESULT 3

004186 PRELIMINARY: PRT: 357 AA.
AC 004186;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Glutamic acid specific protease prepropeptide (EC 3.4.21.19).
OS Staphylococcus aureus.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
ON NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RA Yoshikawa K., Tsuzuki H., Fujiwara T., Nakamura E., Iwamoto H.,
RA Matsumoto K., Shin M., Yoshida N., Teraoka H.;

RT "Purification, characterization and gene cloning of a novel glutamic
acid-specific endopeptidase from staphylococcus aureus atcc 12600.";
RL Biochim. Biophys. Acta 1121:221-228(1991).
DR EMBL: D00730; BAA06030.1; -;
DR MEROPS: S01.269; -;
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000126; Ser_proteas_V8.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00839; V8PROTEASE.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00672; V8_HIS; 1.
DR PROSITE: PS00673; V8_SER; 1.
KW Hydrolyase; Protease; Serine protease.
FT CHAIN 69 357 POTENTIAL.
SQ SEQUENCE 357 AA; 38651 MW; 58AA94AE371E2577 CRC64;

Query Match 100.0%; Score 54; DB 2; Length 357;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPVF 10
|||||
Db 233 TCGNSGSPVF 242

RESULT 4

09AJX0 PRELIMINARY: PRT: 217 AA.
AC 09AJX0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Extracellular serine proteinase precursor (Fragment).
GN ESP.
OS Staphylococcus epidermidis.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
ON NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6746;
RA Dubin G.;
RT "Staphylococcus epidermidis extracellular serine proteinase.";
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ305145; CAC27157.1; -;
DR HSP: P09331; 1EXF.
DR MEROPS: S01.269; -;
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR000126; Ser_proteas_V8.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00839; V8PROTEASE.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00673; V8_SER; 1.
KW Hydrolyase; Serine protease; Signal.
FT NON_TER 1 1
FT SIGNAL 1 1
FT CHAIN 2 217 POTENTIAL.
SQ SEQUENCE 217 AA; 23667 MW; FB9B88BD453BBB7 CRC64;

Query Match 90.7%; Score 49; DB 2; Length 217;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCGSGSPVF 10
|||||
Db 167 GCGSGSPVF 175

RESULT 5

09PC94 PRELIMINARY: PRT: 716 AA.
AC 09PC94;
DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DE Hypothetical protein Xf1887.
OS Xf1887.
GN Xf1887.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=945C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barrios M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briston M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carro D.M., Carter H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nham A.J., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Paixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quagto R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA da Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsunako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AEO4008; AAF84593.1; -;
DR MEROPS: S46.001; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 716 AA; 79375 MW; EAF086E2315BDPC CRC64;
SO
Query Match 88.9%; Score 48; DB 16; Length 716;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCGNSGSPV 9
Db 647 TCGNSGSPV 655
IIIIIIIIII
PRT; 284 AA.
RESULT 6
Q47809 PRELIMINARY;
ID Q47809
AC Q47809
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Staphylococcal serine proteinase homologue.
GN SPRE.
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Enterococcaceae; Enterococcus.
NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OG1-10;
RA Su Y.A., Clewell D.B.;
RT "A gene (spre) downstream of gele of Enterococcus faecalis OG1-10
resembles serine proteinase determinant of Staphylococcus aureus

RT strain V8.";
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z12296; CA478168.1; -;
DR InterPro: IPR001254; Ser_protease_Try.
DR PRINTS: PR00839; V8PROTEASE.
DR PROSITE: PS50240; TRYPsin_DOM; 1.
DR PROSITE: PS00672; V8_HIS; 1.
DR PROSITE: PS00673; V8_SER; UNKNOWN_1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 284 AA; 31063 MW; D4F0312BEE778415 CRC64;
Query Match 81.5%; Score 44; DB 2; Length 284;
Best Local Similarity 70.0%; Pred. No. 7.8;
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
OY 1 TCGNSGSPVF 10
Db 224 TCGNSGSPVF 233
IIIIIIIIII
PRT; 289 AA.
RESULT 7
Q98M08 PRELIMINARY;
ID Q98M08
AC Q98M08
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, last sequence update)
DE Hypothetical protein ms18587.
GN MS18587.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Kanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002995; BAB48055.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 289 AA; 31310 MW; 8947CDPDC6BC0253 CRC64;
SO
Query Match 81.5%; Score 44; DB 16; Length 289;
Best Local Similarity 88.9%; Pred. No. 7.9;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 GGNSGSPVF 10
Db 194 GGNSGSPVF 202
IIIIIIIIII
PRT; 398 AA.
RESULT 8
Q8VE75 PRELIMINARY;
ID Q8VE75
AC Q8VE75
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Similar to RIKEN cDNA 463241K18 gene (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mamalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC019638; AA019638.1; -
FT NON TER 1
SQ SEQUENCE 398 AA; 45268 MW; 51FED8CE693AC1B7 CRC64;

Query Match 81.5%; Score 44; DB 11; Length 398;
Best Local Similarity 88.9%; Pred. No. 11;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGNGSGSPVF 10
11:111111
DB 325 GSGSGSPVF 333

RESULT 9

Q9D2L9 PRELIMINARY; PRT; 613 AA.

AC Q9D2L9;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE 4632417K18RIK protein.
CN 4632417K18RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SKIN;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Alzawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein N.J., Bult C., Fletcher C., Fujita K., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
DR EMBL: AK019499; BAB31763.1; -
DR MGI: MGI:1915508; 4632417K18RIK.
SQ SEQUENCE 613 AA; 69948 MW; F926114F705A639B CRC64;

Query Match 81.5%; Score 44; DB 11; Length 613;
Best Local Similarity 88.9%; Pred. No. 17;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 CGNGSGSPVF 10
11:111111
DB 540 GSGSGSPVF 548

RESULT 10

Q98KC9 PRELIMINARY; PRT; 1008 AA.

AC Q98KC9;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein ml1534.

GN ML1534.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=MAF303099;
RX MEDLINE=21082930; Pubmed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Katanebe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002997; BAB48885.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1008 AA; 100023 MW; AC3F2877D0B94C53 CRC64;

Query Match 79.6%; Score 43; DB 16; Length 1008;
Best Local Similarity 80.0%; Pred. No. 43;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGNGSGSPVF 10
11111111
DB 442 TCGNGSGSTVF 451

RESULT 11

Q8YWM7 PRELIMINARY; PRT; 169 AA.

AC Q8YWM7;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Hypothetical protein ALR5049.
GN ALR5049.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; Pubmed=11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
Katanebe A., Iritiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003598; BAB76748.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 169 AA; 17427 MW; 0A610CC04EB4C48A CRC64;

Query Match 77.8%; Score 42; DB 16; Length 169;
Best Local Similarity 88.9%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGNGSGSPV 9
11111111
DB 109 TCGNSASPV 117

RESULT 12

Q8XHM4 PRELIMINARY; PRT; 458 AA.

AC Q8XHM4;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE UDP-N-acetylmuramoylalanine D-glutamate ligase.
 GN MORD OR CPE2459.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_Taxid=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=13 / TYPE A;
 RX PubMed=11792842;
 RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
 RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.,
 RT "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 DR EMBL: APO03194; BAB82165.1; -
 DR InterPro: IPR000713; Mur_Ligase.
 DR InterPro: IPR004101; Mur_Ligase_C.
 DR Pfam: PF01225; Mur_Ligase; 1.
 DR Pfam: PF02875; Mur_Ligase_C; 1.
 DR TIGRFAMs: TIGR01087; murD; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 458 AA; 51513 MW; CDB19A3365EAE64 CRC64;

Query Match 77.8%; Score 42; DB 16; Length 458;
 Best Local Similarity 70.0%; Pred. No. 28;
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 10
 DB 148 TGGNIGTPTL 157

RESULT 13
 O92DX8 PRELIMINARY; PRT; 315 AA.
 AC O92DX8;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Protease DO (HTRA).
 GN RPI86.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 OX NCBI_Taxid=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MADRID E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Almark U.C.M., Podowski R.M., Neeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.,
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria.";
 RT Nature 396:133-140(1998).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.
 DR EMBL: AJ235270; CAI14652.1; -
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser.protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR PROSITE: PSS0106; PDZ; 1.
 DR PROSITE: PSS0240; TRYPsin_DOM; 1.
 KW Hydrolyase; Serine protease; Complete proteome.
 SQ SEQUENCE 315 AA; 35033 MW; 2D91A054FFBE9A1 CRC64;

Query Match 74.1%; Score 40; DB 16; Length 315;
 Best Local Similarity 88.9%; Pred. No. 44;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9

Db 188 TGGASGSPV 196

RESULT 14
 O62313 PRELIMINARY; PRT; 353 AA.
 AC O62313;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE Trans-golgi network integral membrane protein TGN38A precursor (Trans-
 DE golgi network protein 1) (TGN38 homolog).
 GN TGN1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR: TISSUE=BRAIN;
 RX MEDLINE=95301533; PubMed=7540170;
 RA Kasai K., Takahashi S., Murakami K., Nakayama K.,
 RT "Strain-specific presence of two TGN38 isoforms and absence of TGN41
 RT in mouse.";
 RT J. Biol. Chem. 270:14471-14476(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.,
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND
 CC FROM TRANS-GOLGI NETWORK.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-
 CC GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL
 CC SURFACE RETURNING VIA ENDOSOMES (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
 CC DR EMBL: D50031; BAA08757.1; -
 DR EMBL: BC009143; AAH09143.1; -
 DR MCD: MGI:105080; Tgnt1.
 KW Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.
 FT SIGNAL 1 17
 FT CHAIN 18 353
 FT FT
 FT DOMAIN 18 298
 FT TRANSMEM 299 319
 FT DOMAIN 320 353
 FT DOMAIN 346 349
 FT DOMAIN 131 178
 FT REPEAT 131 138
 FT REPEAT 139 146
 FT REPEAT 147 154
 FT REPEAT 155 162
 FT REPEAT 163 170
 FT REPEAT 171 178
 FT CARBOHYD 110 110
 FT CARBOHYD 293 293
 SQ SEQUENCE 353 AA; 37848 MW; 95C340C2F4A21B3 CRC64;

Query Match 74.1%; Score 40; DB 11; Length 353;
 Best Local Similarity 87.5%; Pred. No. 49;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 8
 DB 138 TGGNSGSPV 145

RESULT 15
 O62314 PRELIMINARY; PRT; 363 AA.
 AC O62314;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
 DE Trans-golgi network integral membrane protein TGN38B precursor (Trans-
 golgi network protein 2) (TGN38 homolog).
 GN TGNL2 OR TGN2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=BRAIN;
 RX MEDLINE=95301533; PubMed=7540170;
 RA Kasai K., Takahashi S., Murakami K., Nakayama K.;
 RT "Strain-specific presence of two TGN38 isoforms and absence of TGN41
 in mouse.";
 RL J. Biol. Chem. 270:14471-14476(1995).
 CC -!- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND
 CC FROM TRANS-GOLGI NETWORK.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-
 CC GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL
 CC SURFACE RETURNING VIA ENDOSOMES (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
 CC -!- MISCELLANEOUS: NOT FOUND IN STRAINS BALB/C, C57BL/6 AND DBA/2.
 DR EMBL: D50032; BAA08758.1; -;
 DR MGD: MGI:105079; Tgoln2.
 KW Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 363 TRANS-GOLGI NETWORK INTEGRAL MEMBRANE
 FT DOMAIN 18 308 PROTEIN TGN38B.
 FT TRANSMEM 309 329 EXTRACELLULAR (POTENTIAL).
 FT SITE 330 363 POTENTIAL.
 FT SITE 356 359 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 133 188 ENDOTOXIN SIGNAL (BY SIMILARITY).
 FT REPEAT 133 140 7 X 8 AA TANDEM REPEATS.
 FT REPEAT 141 148 1.
 FT REPEAT 149 156 2.
 FT REPEAT 147 164 3.
 FT REPEAT 165 172 4.
 FT REPEAT 173 180 5.
 FT REPEAT 181 188 6.
 FT CARBOHYD 303 303 7.
 FT SEQUENCE 363 AA; 38821 MW; 2826FA9E958C5C27 CRC64; N-LINKED (GLCNAC...) (POTENTIAL).

Query Match 74.1%; Score 40; DB 11; Length 363;
 Best Local Similarity 87.5%; Pred. No. 50;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGN38GSP 8
 |||||
 DB 140 TGN38GSP 147

Search completed: December 20, 2002, 12:10:18
 Job time : 63.6842 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:10:29 ; Search time 34.5789 Seconds
(without alignments)
79.820 Million cell updates/sec

Title: US-10-008-355-26

Perfect score: 48

Sequence: 1 TGCNCGSPV 9

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched:

441362 segs, 15338381 residues

Total number of hits satisfying chosen parameters:

882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh
-O=cg2_1/USPRO_POOL/US10008355/runat_17122002_112337_14581/app.query.fasta_1.398
-DB=Issued_Patents_NA -QWTF=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10008355.ecgn1.1.17 @runat.17122002_112337_14581 -NCPU=6 -ICPU=3
-NO_XLPPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: Issued_Patents_NA:*
2: /cg2_6/p2odata/1/ina/5A_COMB.seq:*
3: /cg2_6/p2odata/1/ina/5B_COMB.seq:*
4: /cg2_6/p2odata/1/ina/6A_COMB.seq:*
5: /cg2_6/p2odata/1/ina/6B_COMB.seq:*
6: /cg2_6/p2odata/1/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	1974	4	US-09-221-017B-726
2	48	100.0	2384	4	US-09-221-017B-1045
3	43	89.6	936	4	US-09-134-001C-892
4	41	85.4	758	4	US-09-071-035-427
5	41	85.4	888	4	US-09-071-035-425
6	37	77.1	39	1	US-08-225-224-31
7	37	77.1	39	3	US-08-722-258-31
8	37	77.1	39	5	PCT-US95-04468-31
9	37	77.1	42	1	US-08-225-224-28
10	37	77.1	42	3	US-08-722-258-28
11	37	77.1	42	5	PCT-US95-04468-28
12	37	77.1	743	4	US-08-998-416-856

13	37	77.1	1448	1	US-08-035-634-1	Sequence 1, Appl1
14	36	75.0	454	2	US-08-680-326-118	Sequence 118, App
15	36	75.0	686	4	US-09-221-017B-38	Sequence 38, Appl
16	36	75.0	2774	1	US-07-723-002C-5	Sequence 5, Appl1
17	36	75.0	3958	2	US-07-952-853-21	Sequence 21, Appl
18	36	75.0	3958	2	US-08-914-848-21	Sequence 21, Appl
19	36	75.0	5399	1	US-08-064-121-1	Sequence 1, Appl1
20	36	75.0	5399	1	US-08-478-015-1	Sequence 1, Appl1
21	36	75.0	5399	3	US-08-475-975-1	Sequence 1, Appl1
22	36	75.0	5399	3	US-09-084-889-1	Sequence 1, Appl1
23	36	75.0	6555	1	US-08-351-413-2	Sequence 2, Appl1
24	36	75.0	6555	2	US-09-025-583-2	Sequence 2, Appl1
25	36	75.0	6854	4	US-09-194-905-7	Sequence 7, Appl1
26	36	75.0	7566	2	US-08-232-016-23	Sequence 22, Appl
27	36	75.0	7639	2	US-08-232-016-22	Sequence 22, Appl
28	36	75.0	15397	2	US-08-673-768-1	Sequence 1, Appl1
29	36	75.0	15397	2	US-08-673-768-1	Sequence 1, Appl1
30	36	75.0	24595	6	5428147-1	Patent No. 5428147
31	35	72.9	45	4	US-09-091-814-104	Sequence 104, App
32	35	72.9	55	4	US-09-091-814-109	Sequence 109, App
33	35	72.9	62	4	US-09-364-539-58	Sequence 58, Appl
34	35	72.9	62	4	US-09-364-539-59	Sequence 59, Appl
35	35	72.9	423	4	US-09-342-681C-94	Sequence 94, Appl
36	35	72.9	433	1	US-08-549-757A-13	Sequence 13, Appl1
37	35	72.9	663	1	US-07-624-313-3	Sequence 290, Appl
38	35	72.9	700	4	US-09-328-111-290	Sequence 53, Appl
39	35	72.9	699	4	US-09-171-209-53	Sequence 262, App
40	35	72.9	738	4	US-08-998-416-262	Sequence 38, Appl1
41	35	72.9	738	4	US-09-182-145-38	Sequence 1, Appl1
42	35	72.9	830	3	US-09-165-240-1	Sequence 39, Appl
43	35	72.9	830	4	US-09-568-059-1	Sequence 1, Appl1
44	35	72.9	841	4	US-09-182-145-39	Sequence 39, Appl
45	35	72.9	917	4	US-09-227-357-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-09-221-017B-726
; Sequence 726, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998

No Patent

```
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 726:
SEQUENCE CHARACTERISTICS:
LENGTH: 1974 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...1974
US-09-221-017B-726

Alignment Scores:
Pred. No.: 7.9      Length: 1974
Score: 48.00      Matches: 9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 4      Gaps: 0

US-10-008-355-26 (1-9) x US-09-221-017B-726 (1-1974)

QY 1 ThnGlyGlyAsnSerGlySerProval 9
DB 1524 ACCGGCGGTACTCGGTAGCCCGGTA 1550

RESULT 2
US-09-221-017B-1045
Sequence 1045, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: ROSS, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FastSeq for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
```

```
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: MONROY, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEFAX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 1045:
SEQUENCE CHARACTERISTICS:
LENGTH: 2384 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: PORYPHYROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...2384
US-09-221-017B-1045

Alignment Scores:
Pred. No.: 9.68      Length: 2384
Score: 48.00      Matches: 9
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 4      Gaps: 0

US-10-008-355-26 (1-9) x US-09-221-017B-1045 (1-2384)

QY 1 ThnGlyGlyAsnSerGlySerProval 9
DB 420 ACCGGCGGCACTCAGCGAGTCCGGTC 446

RESULT 3
US-09-134-001C-892
Sequence 892, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 892
LENGTH: 936
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-892

Alignment Scores:
Pred. No.: 28      Length: 936
Score: 43.00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 89.58%      Indels: 0
DB: 4      Gaps: 0

US-10-008-355-26 (1-9) x US-09-134-001C-892 (1-936)

QY 2 GlyGlyAsnSerGlySerProval 9
|||||
```

Db 781 GGtGAAACTCTGATCTCCAGTA 804

RESULT 4
US-09-071-035-427
Sequence 427, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 427:
SEQUENCE CHARACTERISTICS:
LENGTH: 758 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-09-071-035-427

Alignment Scores:
Pred. No.: 51
Score: 41.00
Percent Similarity: 88.89%
Best Local Similarity: 77.78%
Query Match: 85.42%
DB: 4
Gaps: 0

US-10-008-355-26 (1-9) x US-09-071-035-427 (1-758)

QY 1 ThrglyGlyanserGlySerProval 9
||||||| |||||||||
Db 575 ACCGGCGTCATCTGTTCCACCATC 601

RESULT 5
US-09-071-035-425
Sequence 425, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071,035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 425:
SEQUENCE CHARACTERISTICS:
LENGTH: 888 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-09-071-035-425

Alignment Scores:
Pred. No.: 60.5
Score: 41.00
Percent Similarity: 88.89%
Best Local Similarity: 77.78%
Query Match: 85.42%
DB: 4
Gaps: 0

US-10-008-355-26 (1-9) x US-09-071-035-425 (1-888)

QY 1 ThrglyGlyanserGlySerProval 9
||||||| |||||||||
Db 703 ACCGGCGTCATCTGTTCCACCATC 729

RESULT 6
US-08-225-224-31/C
Sequence 31, Application US/08225224
Patent No. 5635599
GENERAL INFORMATION:
APPLICANT: PASTAN, Ira
TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND
TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105-1493

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,224
FILING DATE: 8-APR-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen L.
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 15280-193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600

TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (primer)
US-08-225-224-31

Alignment Scores:
Pred. No.: 10.9 Length: 39
Score: 37.00 Matches: 6
Percent Similarity: 87.50% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 77.08% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-26 (1-9) x US-08-225-224-31 (1-39)

QY 1 ThGlyGlyAsnSerglyserPro 8
|||||
24 ACCGAGGTAACGCTGGGCACCT 1

RESULT 7
US-08-722-258-31/c
; Sequence 31, Application US/08722258
; Patent No. 6011002
; GENERAL INFORMATION:
; APPLICANT: Pastan, Ira
; APPLICANT: Kreitman, Robert J.
; APPLICANT: Puri, Raj K.
; TITLE OF INVENTION: Circularly Permuted Ligands and
; TITLE OF INVENTION: Circularly Permuted Chimeric Molecules
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-5834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,258
; FILING DATE: 08-JAN-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/04468
; FILING DATE: 06-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/225,224
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 015280-193100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -

LOCATION: 1..39
; OTHER INFORMATION: /note="BK-138 primer"
US-08-722-258-31

Alignment Scores:
Pred. No.: 10.9 Length: 39
Score: 37.00 Matches: 6
Percent Similarity: 87.50% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 77.08% Indels: 0
DB: 3 Gaps: 0

US-10-008-355-26 (1-9) x US-08-722-258-31 (1-39)

QY 1 ThGlyGlyAsnSerglyserPro 8
|||||
24 ACCGAGGTAACGCTGGGCACCT 1

RESULT 8
PCT-US95-04468-31/c
; Sequence 31, Application PC/TUS9504468
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND
; TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS
; NUMBER OF SEQUENCES: 59
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04468
; FILING DATE: 07-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/225,224
; FILING DATE: 08-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-193-1PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 31:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (primer)
PCT-US95-04468-31

Alignment Scores:
Pred. No.: 10.9 Length: 39
Score: 37.00 Matches: 6
Percent Similarity: 87.50% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 77.08% Indels: 0
DB: 5 Gaps: 0

US-10-008-355-26 (1-9) x PCT-US95-04468-31 (1-39)

QY 1 ThGlyGlyAsnSerglyserPro 8
|||||
24 ACCGAGGTAACGCTGGGCACCT 1

RESULT 9
US-08-225-224-28
; Sequence 28, Application US/08225224
; Patent No. 5635599
; GENERAL INFORMATION:

APPLICANT: PASTAN, Ira
APPLICANT: KREITMAN, Robert J.
TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND
TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: Stewart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/225,224
FILING DATE: 8-APR-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen L.
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 15280-193
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (primer)
US-08-225-224-28

Alignment Scores:
Pred. No.: 11.8 Length: 42
Score: 37.00 Matches: 6
Percent Similarity: 87.50% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 77.08% Indels: 0
Gaps: 0
DB: 1

US-10-008-355-26 (1-9) x US-08-225-224-28 (1-42)

QY 1 ThGlyGlyAsnSerGlySerPro 8
|||||
Db 7 ACCGAGGTAAAGTGGGCACCT 30

RESULT 10
US-08-722-258-28
Sequence 28, Application US/08722258
GENERAL INFORMATION:
APPLICANT: Pastan, Ira
APPLICANT: Kreitman, Robert J.
TITLE OF INVENTION: Circularly Permuted Ligands and
TITLE OF INVENTION: Circularly Permuted Chimeric Molecules
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,258
FILING DATE: 08-JAN-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/04468
FILING DATE: 06-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/225,224
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 015280-193100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0300
TELEFAX: (415) 576-0200
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 42 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: -
LOCATION: 1..42
OTHER INFORMATION: /note="BK-135 primer"
US-08-722-258-28

Alignment Scores:
Pred. No.: 11.8 Length: 42
Score: 37.00 Matches: 6
Percent Similarity: 87.50% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 77.08% Indels: 0
Gaps: 0
DB: 3

US-10-008-355-26 (1-9) x US-08-722-258-28 (1-42)

QY 1 ThGlyGlyAsnSerGlySerPro 8
|||||
Db 7 ACCGAGGTAAAGTGGGCACCT 30

RESULT 11
PCT-US95-04468-28
Sequence 28, Application PC/TUS9504468
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: CIRCULARLY PERMUTATED LIGANDS AND
TITLE OF INVENTION: CIRCULARLY PERMUTATED FUSION PROTEINS
NUMBER OF SEQUENCES: 59
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04468
FILING DATE: 07-APR-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/225,224
FILING DATE: 08-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen L.
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 15280-193-1PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 28:

SEQUENCE CHARACTERISTICS:
LENGTH: 42 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (primer)
PCT-US95-04468-28

Alignment Scores:
Pred. No.: 11.8 Length: 42
Score: 37.00 Matches: 6
Percent Similarity: 87.50% Conservative: 1
Best Local Similarity: 75.00% Mismatches: 1
Query Match: 77.08% Indels: 0
DB: Gaps: 0

US-10-008-355-26 (1-9) x PCT-US95-04468-28 (1-42)

QY 1 ThrglyGlyAsnSerGlySerPro 8
|||||
Db 7 ACCGAGGTACCGTGGCGACCT 30

RESULT 12
US-08-998-416-856/C
Sequence 856, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philppsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 856:
SEQUENCE CHARACTERISTICS:
LENGTH: 743 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1539UP
US-08-998-416-856

Alignment Scores:
Pred. No.: 262 Length: 743
Score: 37.00 Matches: 6
Percent Similarity: 77.78% Conservative: 1
Best Local Similarity: 66.67% Mismatches: 2
Query Match: 77.08% Indels: 0
DB: Gaps: 0

US-10-008-355-26 (1-9) x US-08-998-416-856 (1-743)

QY 1 ThrglyGlyAsnSerGlySerProval 9
|||||
Db 576 ACCGGTGTGATCTGAGACCTATA 550

RESULT 13
US-08-035-634-1
Sequence 1, Application US/08035634
Patent No. 5459064
GENERAL INFORMATION:
APPLICANT: SHIONOGI & CO., LTD.
TITLE OF INVENTION: A NO. 5459064el Protease
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: EPSON PC-286 Book type laptop
OPERATING SYSTEM: MS-DOS 2.11
SOFTWARE: Wordstar 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/035,634
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/782,372
FILING DATE: 24-OCT-1991
APPLICATION NUMBER: Japanese Patent
APPLICATION NUMBER: Application No. 5459064 2-288110
FILING DATE: October 24, 1990
ATTORNEY/AGENT INFORMATION:
NAME: BOZICEVIC, KARL
REGISTRATION NUMBER: 28,807
REFERENCE/DOCKET NUMBER: 29900-20298.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 677-7000
TELEFAX: (415) 677-7522
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1448 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Bacillus licheniformis
STRAIN: ATCC NO. 14580

FEATURE:
NAME/KEY: coding sequence
LOCATION: 323 to 1270
IDENTIFICATION METHOD: by experiment
NAME/KEY: signal peptide
LOCATION: 323 to 604
IDENTIFICATION METHOD: by experiment
NAME/KEY: mature peptide
LOCATION: 605 to 1270
IDENTIFICATION METHOD: by experiment
OTHER INFORMATION:

DB: 4 Gaps: 0

US-10-008-355-26 (1-9) x US-09-221-017B-38 (1-686)

QY 1 ThrGlyGlyAsnSerGlySerProVal 9

Db 482 ACGGCGGATCATCCGCTTCCCGATC 456

Search completed: December 20, 2002, 14:11:52
Job time : 36.5789 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:17:34 ; Search time 36 Seconds
(without alignments)
99.214 Million cell updates/sec

Title: US-10-008-355-26
Perfect score: 48
Sequence: 1 TCGNSGSPV 9

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 356696 seqs, 198428768 residues
Total number of hits satisfying chosen parameters: 713392

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n model -DEV=xlh
-Q/cgcn2_1/USPTO_spool/US10008355/runat_17122002_112339_14635/app.query.fasta_1.398
-DB=published_Applications_NA -QMT=fastlap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10008355 @CGN_1.1.21 @runat_17122002_112339_14635
-NUP=6 -ICPR=3 -NO_XLPEXY -NO_MAP -LARGEOUTERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

1: /cgcn2_6/ptodata/2/pubpna/US07_PUBSCOMB.seq:*
2: /cgcn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq:*
3: /cgcn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgcn2_6/ptodata/2/pubpna/US06_PUBSCOMB.seq:*
5: /cgcn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgcn2_6/ptodata/2/pubpna/PTCTUS_PUBSCOMB.seq:*
7: /cgcn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgcn2_6/ptodata/2/pubpna/US08_PUBSCOMB.seq:*
9: /cgcn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgcn2_6/ptodata/2/pubpna/US09_PUBSCOMB.seq:*
11: /cgcn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgcn2_6/ptodata/2/pubpna/US10_PUBSCOMB.seq:*
13: /cgcn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgcn2_6/ptodata/2/pubpna/US60_PUBSCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	2139	9	US-10-008-355-1
2	41	85.4	4951	10	US-09-070-927A-261
3	40	83.3	450	10	US-09-867-701-1705
4	40	83.3	7441	10	US-09-764-878-243

5	39	81.2	1775	9	US-10-001-887-75	Sequence 75, Appl
6	38	79.2	454	10	US-09-770-444-591	Sequence 591, App
7	38	79.2	491	10	US-09-783-550-6501	Sequence 6501, Ap
8	38	79.2	1114	10	US-09-808-701-14	Sequence 14, Appl
9	37.5	78.1	400	10	US-09-960-352-14943	Sequence 14943, A
10	37	77.1	222	10	US-09-864-761-32393	Sequence 32393, A
11	37	77.1	534	10	US-09-974-300-327	Sequence 327, App
12	37	77.1	557	10	US-09-864-761-15889	Sequence 15889, A
13	37	77.1	720	10	US-09-815-242-4296	Sequence 4296, Ap
14	37	77.1	765	10	US-09-815-242-8179	Sequence 8179, Ap
15	37	77.1	2000	9	US-09-938-842A-3053	Sequence 3053, Ap
16	37	77.1	2205	10	US-09-822-830A-458	Sequence 458, App
17	37	77.1	51552	10	US-09-733-294A-30	Sequence 30, Appl
18	36	75.0	191	10	US-09-864-761-24491	Sequence 24491, A
19	36	75.0	297	10	US-09-783-590-10996	Sequence 10996, A
20	36	75.0	393	10	US-09-960-332-3735	Sequence 3735, Ap
21	36	75.0	441	10	US-09-960-352-9622	Sequence 9622, Ap
22	36	75.0	529	10	US-09-864-761-7787	Sequence 7787, Ap
23	36	75.0	1142	9	US-09-764-868-45	Sequence 45, Appl
24	36	75.0	1374	10	US-09-070-927A-650	Sequence 650, App
25	36	75.0	2370	9	US-09-712-363-98	Sequence 98, Appl
26	36	75.0	2667	9	US-09-938-842A-1308	Sequence 1308, Ap
27	36	75.0	2764	10	US-09-925-301-326	Sequence 326, App
28	36	75.0	4947	10	US-09-118-276-21	Sequence 21, Appl
29	36	75.0	17397	10	US-09-764-869-1945	Sequence 1945, Ap
30	36	75.0	19334	10	US-09-764-869-1943	Sequence 1943, Ap
31	36	75.0	19345	10	US-09-764-869-1944	Sequence 1944, Ap
32	36	75.0	30625	10	US-09-927-091-5	Sequence 5, Appl1
33	35	72.9	192	10	US-09-864-761-31758	Sequence 31758, A
34	35	72.9	201	9	US-10-046-935-406	Sequence 406, App
35	35	72.9	201	9	US-09-878-178-406	Sequence 406, App
36	35	72.9	218	10	US-09-815-343-1415	Sequence 1415, Ap
37	35	72.9	222	10	US-09-815-343-676	Sequence 676, App
38	35	72.9	222	10	US-09-815-343-1285	Sequence 1285, App
39	35	72.9	240	10	US-09-864-761-21097	Sequence 21097, A
40	35	72.9	243	10	US-09-815-343-614	Sequence 614, App
41	35	72.9	247	10	US-09-864-761-29769	Sequence 29769, A
42	35	72.9	253	10	US-09-815-343-729	Sequence 729, App
43	35	72.9	258	10	US-09-923-876-989	Sequence 989, App
44	35	72.9	289	10	US-09-923-876-4593	Sequence 4593, Ap
45	35	72.9	299	10	US-09-998-598-618	Sequence 618, App

ALIGNMENTS

RESULT 1
US-10-008-355-1
; Sequence 1, Application US/10008355
; Patent No. US20020164759A1
GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; PRIORITY FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIORITY FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-1

Alignment Scores:
Pred. No.: 4 69
Score: 48.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Length: 2139
Matches: 9
Conservative: 0
Mismatches: 0
Indels: 0

DB: 9 Gaps: 0

US-10-008-355-26 (1-9) x US-10-008-355-1 (1-2139)

QY 1 ThnglyGlyAsnSerGlySerProval 9

DB 1930 ACGGCGGTAACTCGGTAGCCCGCA 1956

RESULT 2

US-09-070-927A-261/C

Sequence 261, Application US/09070927A

Patent No. US20020120116A1

GENERAL INFORMATION:

APPLICANT: Charles A. Kunsch

Patrick J. Dillon

Steven Barash

TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides

NUMBER OF SEQUENCES: 982

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/046,655

FILING DATE: 1997-05-16

APPLICATION NUMBER: 60/044,031

FILING DATE: 1997-05-06

APPLICATION NUMBER: 60/066,009

FILING DATE: 1997-11-14

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB369

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 261:

SEQUENCE CHARACTERISTICS:

LENGTH: 4951 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 261:

US-09-070-927A-261

Alignment Scores:

Pred. No.: 226 Length: 4951

Score: 41.00 Matches: 7

Percent Similarity: 88.89% Conservative: 1

Best Local Similarity: 77.78% Mismatches: 1

Query Match: 85.42% Indels: 0

DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x US-09-070-927A-261 (1-4951)

QY 1 ThnglyGlyAsnSerGlySerProval 9

DB 695 ACCGGCGTCAATCTGTCACCAATC 669

RESULT 3

US-09-867-701-1705

Sequence 1705, Application US/09867701

Patent No. US20020132237A1

GENERAL INFORMATION:

APPLICANT: Aglate, Paul A.

APPLICANT: Jones, Robert

APPLICANT: Harlocker, Susan L.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

FILE REFERENCE: 210121.497

CURRENT APPLICATION NUMBER: US/09/867,701

CURRENT FILING DATE: 2001-03-29

NUMBER OF SEQ ID NOS: 10912

SOFTWARE: FastSeq for Windows version 4.0

SEQ ID NO 1705

LENGTH: 450

TYPE: DNA

ORGANISM: Homo sapien

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(450)

OTHER INFORMATION: n = A,T,C or G

US-09-867-701-1705

Alignment Scores:

Pred. No.: 25.5 Length: 450

Score: 40.00 Matches: 7

Percent Similarity: 87.50% Conservative: 0

Best Local Similarity: 83.33% Mismatches: 1

Query Match: 83.33% Indels: 0

DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x US-09-867-701-1705 (1-450)

QY 1 ThnglyGlyAsnSerGlySerPro 8

DB 82 ACAGGAGGTAAATTCGCAAGCC 105

RESULT 4

US-09-764-878-243/C

Sequence 243, Application US/09764878

Patent No. US20020090615A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PA121

CURRENT APPLICATION NUMBER: US/09/764,878

CURRENT FILING DATE: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 428

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 243

LENGTH: 7441

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-878-243

Alignment Scores:

Pred. No.: 537 Length: 7441

Score: 40.00 Matches: 7

Percent Similarity: 100.00% Conservative: 1

Best Local Similarity: 87.50% Mismatches: 0

Query Match: 83.33% Indels: 0

DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x US-09-764-878-243 (1-7441)

QY 2 GlyGlyAsnSerGlySerProval 9

DB 2023 GCGGTAAACACAGCAGCCCTGTCG 2000

RESULT 5

US-10-001-887-75

Sequence 75, Application US/10001887

```
; Patent No. US20020155464A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herive
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Genes and PR
; FILE REFERENCE: DEX-0269
; CURRENT APPLICATION NUMBER: US/10/001,887
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,998
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: 60/252,563
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 75
; LENGTH: 1775
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-887-75

Alignment Scores:
Pred. No.: 173          Length: 1775
Score: 39.00           Matches: 7
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 81.25%          Indels: 0
DB: 9                   Gaps: 0

US-10-008-355-26 (1-9) x US-10-001-887-75 (1-1775)
QY 2 GtGtGtGtCAAGTGTCTCCGTC 8
Db 396 GGGGGGAATTCAGGCTCACCACCA 416

RESULT 6
US-09-770-444-591
; Sequence 591, Application US/09770444
; Patent No. US2002023280A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: Ah, Yong-Olang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Kricker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2027 (PARA-016PRV)
; CURRENT APPLICATION NUMBER: US/09/770,444
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,502
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 591
; LENGTH: 454
```

```
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-770-444-591

Alignment Scores:
Pred. No.: 59.9          Length: 454
Score: 38.00           Matches: 7
Percent Similarity: 100.00%  Conservative: 1
Best Local Similarity: 87.50%  Mismatches: 0
Query Match: 79.17%          Indels: 0
DB: 10                   Gaps: 0

US-10-008-355-26 (1-9) x US-09-770-444-591 (1-454)
QY 2 GtGtGtGtCAAGTGTCTCCGTC 9
Db 202 GGTGTTCAGTGTCTCCGTC 225

RESULT 7
US-09-783-590-6501/c
; Sequence 6501, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16,261
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6501
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (110)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (291)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (356)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (397)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (413)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (421)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (428)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (458)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (463)
; OTHER INFORMATION: n equals a,t,g, or c
```

```
; NAME/KEY: misc feature
; LOCATION: (464)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-6501

Alignment Scores:
Pred. No.: 65.2      Length: 491
Score: 38.00        Matches: 7
Percent Similarity: 100.00%  Conservative: 1
Best Local Similarity: 87.50%  Mismatches: 0
Query Match: 79.17%  Indels: 0
DB: 10              Gaps: 0

US-10-008-355-26 (1-9) x US-09-783-590-6501 (1-491)

QY 2 ThGlyGlyAsnSerGlySerProval 9
   |||||:::|||||
Db 215 GGtGGtCAAGtGGtTCACCCGtG 192

RESULT 8
US-09-808-701-14
; Sequence 14, Application US/09808701
; Patent No. US20020146757A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Zhou, Ping
; APPLICANT: Goodrich, Ryle
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Xue, Aidong J.
; APPLICANT: Ren, Feiyun
; APPLICANT: Wang, Dunrui
; APPLICANT: Chen, Rui-hong
; APPLICANT: Drmanac, Radomir T.
; TITLE OF INVENTION: NO. US20020146757A1el Nucleic Acids and
; FILE REFERENCE: 790CIP2D
; CURRENT APPLICATION NUMBER: US/09/808,701
; CURRENT FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: pt_FL-genes Version 2.0
; SEQ ID NO 14
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (131)..(952)
US-09-808-701-14

Alignment Scores:
Pred. No.: 159      Length: 1114
Score: 38.00        Matches: 7
Percent Similarity: 77.78%  Conservative: 0
Best Local Similarity: 77.78%  Mismatches: 2
Query Match: 79.17%  Indels: 0
DB: 10              Gaps: 0

US-10-008-355-26 (1-9) x US-09-808-701-14 (1-1114)

QY 1 ThGlyGlyAsnSerGlySerProval 9
   |||||
Db 526 ACCGGGGAGACGAGGGCTCCACAGTA 552

RESULT 9
US-09-960-352-14943
; Sequence 14943, Application US/09960352
; Patent No. US20020137139A1

; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathalagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 14943
; LENGTH: 400
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 64-LIB2809-023-Q1-E1-H12
US-09-960-352-14943

Alignment Scores:
Pred. No.: 64.5      Length: 400
Score: 37.50        Matches: 9
Percent Similarity: 90.00%  Conservative: 0
Best Local Similarity: 90.00%  Mismatches: 0
Query Match: 78.12%  Indels: 1
DB: 10              Gaps: 1

US-10-008-355-26 (1-9) x US-09-960-352-14943 (1-400)

QY 1 ThGlyGlyAsn---SerGlySerProval 9
   |||||
Db 110 ACAGCGGTAACCTTCAGGGTCCACAGT 139

RESULT 10
US-09-864-761-32393/C
; Sequence 32393, Application US/09864761
; Patent No. US2002048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
```

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 32393
LENGTH: 222
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121716.16
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
OTHER INFORMATION: SWISSPROT HIT: O52325, EVALUATE 3.70e+00
OTHER INFORMATION: NT HIT: AB029040.1, EVALUATE 1.00e-122
OTHER INFORMATION: EST_HUMAN HIT: BE002805.1, EVALUATE 1.00e-122
US-09-864-761-32393

Alignment Scores:

Pred. No.: 42 Length: 222
Score: 37.00 Matches: 6
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 77.08% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x US-09-864-761-32393 (1-222)

QY 2 GYGLYAsnSerGlySerProval 9

Db 69 GGAGGACACTCAGGAGTCTATC 46

RESULT 11

US-09-974-300-327
Sequence 327, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth
TITLE OF INVENTION: Methods for Monitoring Multiple Gene
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 327
LENGTH: 534
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-327

Alignment Scores:

Pred. No.: 109 Length: 534
Score: 37.00 Matches: 7
Percent Similarity: 87.50% Conservative: 0
Best Local Similarity: 87.50% Mismatches: 1
Query Match: 77.08% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x US-09-974-300-327 (1-534)

QY 2 GYGLYAsnSerGlySerProval 9

Db 69 GGAGGACACTCAGGAGTCTATC 46

Db 357 GGAGGACAAAGCGTTCACCGGTA 380

RESULT 12

US-09-864-761-15889/c
Sequence 15889, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aeomica-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 15889
LENGTH: 557
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL121716.16
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
US-09-864-761-15889

Alignment Scores:

Pred. No.: 114 Length: 557
Score: 37.00 Matches: 6
Percent Similarity: 100.00% Conservative: 2
Best Local Similarity: 75.00% Mismatches: 0
Query Match: 77.08% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x US-09-864-761-15889 (1-557)

Oy 2 GlyAsnSerGlySerProVal 9
|||||:|||||:|||||:
Db 167 GGAGACACTGATCAGAGTCTATC 144

RESULT 13
US-09-815-242-4296
; Sequence 4296, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4296
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4296

Alignment Scores:
Pred. No.: 151 Length: 720
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.08% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x US-09-815-242-4296 (1-720)

Oy 3 GlyAsnSerGlySerProVal 9
|||||:|||||:|||||:
Db 571 GGAACTCTGATCACCCTGTA 591

RESULT 14
US-09-815-242-8179
; Sequence 8179, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8179
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(765)
US-09-815-242-8179

FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8179
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(765)
US-09-815-242-8179

Alignment Scores:
Pred. No.: 161 Length: 765
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 77.08% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x US-09-815-242-8179 (1-765)

Oy 3 GlyAsnSerGlySerProVal 9
|||||:|||||:|||||:
Db 613 GGAACTCTGATCACCCTGTA 633

RESULT 15
US-09-938-842A-3053
; Sequence 3053, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPT300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3053
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3053

Alignment Scores:
Pred. No.: 458 Length: 2000
Score: 37.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00%
 Query Match: 77.08%
 DB: 9
 Mismatches: 0
 Indels: 0
 Gaps: 0

US-10-008-355-26 (1-9) x US-09-938-842A-3053 (1-2000)

QY 1 ThrGlyGlyAsnSerGlySer 7
 |||||
 DB 1897 ACCGAGGTAATCTGGATCT 1917

Search completed: December 20, 2002, 17:10:43
 Job time : 40 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:02:33 ; Search time 171.474 Seconds

(without alignments)
118.199 Million cell updates/sec

Title: US-10-008-355-26

Perfect score: 48

Sequence: 1 TGGNGSPV 9

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame.p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO/US10008355/runat_17122002_112336_14573/app_query.fasta_1.398
-DB=N.Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pro -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10008355 -ECGN1_1.0 -runat_17122002_112336_14573 -NCPU=6 -ICPU=3
-NO_XLPEXT -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECUT=120
-WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N.Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	48	100.0	1019	18	AAV75063	Staphylococcus aur
2	48	100.0	1558	13	AAV27988	Protease from S. A
3	48	100.0	1586	13	AAQ27987	Protease from S. A
4	48	100.0	2139	24	AAI43635	Porphyromonas ging
5	43	89.6	849	22	AAH52485	S. epidermidis ope
6	43	89.6	936	24	ABN91429	Staphylococcus epi
7	43	89.6	3189	22	AAH54330	S. epidermidis gen
8	41	85.4	758	20	AAK20213	Enterococcus faeca
9	41	85.4	758	24	ABN98198	E faecalis EF110 g
10	41	85.4	888	20	AAK20212	Enterococcus faeca
11	41	85.4	888	24	ABN98197	E faecalis EF110 g
12	41	85.4	2130	24	ABQ91212	M. capsulatus gene
13	41	85.4	4951	20	AAK13198	Enterococcus faeca
14	40	83.3	450	24	ABL78727	Human ovarian canc
15	40	83.3	1127	21	AAV07536	Fusarium venenatum
16	40	83.3	7441	22	AAK29979	Human Lung Antigen
17	40	83.3	44861	24	AAK20000	DNA encoding pyrid
18	39	81.2	152	22	AAV79971	Nucleotide sequenc
19	39	81.2	473	23	ABV56742	Human prostate exp
20	39	81.2	741	21	AAV14488	Aspergillus oryzae
21	39	81.2	1173	24	ABK32891	DNA encoding C. al
22	39	81.2	1261	20	AAK04382	Human secreted pro
23	39	81.2	1303	22	AAV26291	Pseudomonas sp typ
24	39	81.2	1546	24	ABQ54166	Human ovarian anti
25	39	81.2	1775	24	ABT03070	Human breast speci
26	39	81.2	2012	21	AAK05942	Human secreted pro
27	39	81.2	2012	22	AAH33087	Human colon cancer
28	39	81.2	2063	24	ABA96872	Human Flaprotein
29	39	81.2	3820	23	ABU06881	Drosophila melanog
30	39	81.2	8578	23	ABU06880	Drosophila melanog
31	38	79.2	402	22	AAI17622	Human breast cance
32	38	79.2	423	22	AAI00017	Human reproductive
33	38	79.2	448	21	AAK27896	Human secreted pro
34	38	79.2	454	24	ABU93826	Arabidopsis thalia
35	38	79.2	828	9	AAAN0912	Sequence encoding
36	38	79.2	828	21	AAA07168	Pig lung protease
37	38	79.2	994	22	AAK59888	Human novel cytol
38	38	79.2	1005	23	ABL22883	Drosophila melanog
39	38	79.2	1114	22	AAK59845	Human novel cytol
40	38	79.2	1119	22	AAH68579	Human protein HP10
41	38	79.2	1719	22	AAK70665	Human immune/haema
42	38	79.2	1723	22	AAK70664	Human immune/haema
43	38	79.2	2421	20	AAV80633	Kidney injury asso
44	38	79.2	2685	22	AAH14840	Human CDNA sequenc
45	38	79.2	2690	23	ABL22872	Drosophila melanog

ALIGNMENTS

RESULT 1
AAV75063
ID AAV75063 standard; DNA; 1019 BP.

XX AAV75063;

AC AAV75063;

DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #752.

XX Computer readable medium; vaccine; S. aureus infection; immunodetection;

XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

XX skin infection; surgical wound infection; scalded skin syndrome;

XX toxic shock syndrome; ds.

XX Staphylococcus aureus.

OS Staphylococcus aureus.

XX Staphylococcus aureus.

XX Staphylococcus aureus.

FH Key Location/Qualifiers

FT misc-feature 901..960
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
XX
XX EP786519-A2.
XX
XX 30-JUL-1997.
XX
XX
XX 07-JAN-1997; 97EP-0100117.
XX
XX 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
XX Rosen CA;
XX
XX WPI; 1997-374922/35.
XX
XX Polynucleotide(s) and proteins derived from Staphylococcus aureus
XX scored on computer readable medium and used in the production of
XX anti-S. aureus vaccines
XX
XX Claim 1; Page 1646-1647; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX of the invention. The DNA sequences are recorded on a computer readable
XX medium, preferably selected from a floppy or hard disk, random access
XX memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX the S. aureus DNA sequences allows putative functions to be assigned so
XX that protein-encoding or regulatory regions of commercial, therapeutic or
XX industrial importance can be obtained. Specifically, sequences which are
XX likely to encode antigens have been identified and these polypeptides can
XX be used in a vaccine composition against S. aureus infection. The
XX polypeptides can also be used in a kit for the immunodetection of
XX S. aureus in a sample. S. aureus is implicated in numerous human diseases,
XX including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX skin and surgical wound infections, scalded skin syndrome, toxic shock
XX syndrome, etc. Organisms transformed with the DNA sequences can be used
XX for recombinant production of the polypeptides. The new DNA sequences
XX (and their fragments) are useful as primers or probes for isolating
XX CC homologues of any of the S. aureus DNA sequences contained on the
XX computer readable medium.
XX
XX SQ Sequence 1019 BP; 364 A; 194 C; 159 G; 240 T; 62 other;
XX
XX Alignment Scores:
XX Pred. No.: 21.5 Length: 1019
XX Score: 48.00 Matches: 9
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 18 Gaps: 0
XX
XX US-10-008-355-26 (1-9) x AAV75063 (1-1019)
XX
XX QY 1 ThGlyGlyAsnSerGlySerProVal 9
XX ||||||||||||||||||||||||||||
XX DB 642 ACTGCTGTAATTCAGGTCACCTGTA 668
XX
XX RESULT 2
XX AAQ27988
XX ID AAQ27988 standard; DNA; 1558 BP.
XX
XX AC AAQ27988;
XX
XX DT 11-FEB-1993 (first entry)
XX
XX DE Protease from S. Aureus.
XX
XX

KW Protease; PCR; amplify; Staphylococcus; ss.
XX
XX Staphylococcus aureus.
OS
XX
XX Key Location/Qualifiers
XX CDS 354..1364
XX FT /*tag= a
XX sig_peptide 354..557
XX FT /*tag= b
XX mat_peptide 558..1361
XX FT /*tag= c
XX
XX JP04211370-A.
XX
XX PD 03-AUG-1992.
XX
XX PR 19-FEB-1991; 91JP-0024633.
XX
XX PR 20-FEB-1990; 90JP-0040398.
XX
XX PA (SHIO) SHIONOGI & CO LTD.
XX
XX WPI; 1992-304938/37.
XX
XX DR P-PSDB; AAR29644.
XX
XX PT Novel protease prep'd. using Bacillus or Saccharomyces host
XX capable of cleaving peptide bond at carboxyl terminus of glutamic
XX acid residues in polypeptide(s)
XX
XX PS Disclosure; Page 15-16; 25pp; Japanese.
XX
XX The sequences given in AAQ27987-88 encode proteases which were isolated
XX from Staphylococcus aureus strains. The DNA sequences were isolated
XX by PCR using the primer sequences given in AAQ27960-86. The protease
XX specifically cleaves the peptide bond at the C-terminus of the
XX glutamic acid residue in polypeptide.
XX
XX SQ Sequence 1558 BP; 579 A; 282 C; 221 G; 476 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 33.8 Length: 1558
XX Score: 48.00 Matches: 9
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 100.00% Indels: 0
XX DB: 13 Gaps: 0
XX
XX US-10-008-355-26 (1-9) x AAQ27988 (1-1558)
XX
XX QY 1 ThGlyGlyAsnSerGlySerProVal 9
XX ||||||||||||||||||||||||||||
XX DB 1050 ACTGCTGTAATTCAGGTCACCTGTA 1076
XX
XX RESULT 3
XX AAQ27987
XX ID AAQ27987 standard; DNA; 1586 BP.
XX
XX AC AAQ27987;
XX
XX DT 11-FEB-1993 (first entry)
XX
XX DE Protease from S. Aureus ATCC12600.
XX
XX KW Protease; PCR; amplify; Staphylococcus; ss.
XX
XX OS Staphylococcus aureus.
XX
XX Key Location/Qualifiers
XX CDS 352..1425
XX FT /*tag= a
XX sig_peptide 352..555
XX FT /*tag= b
XX mat_peptide 556..1425
XX

FT /*tag= C
XX JP04211370-A.
XX
XX 03-AUG-1992.
XX
XX 19-FEB-1991; 91JP-0024633.
XX
XX 20-FEB-1990; 90JP-0040398.
XX
XX (SHIO) SHIONOGI & CO LTD.
XX
XX WPI: 1992-304938/37.
XX P-PSDB; AAR26842.
XX
XX Novel protease prepd. using Bacillus or Saccharomyces host -
PT capable of cleaving peptide bond at carboxyl terminus of glutamic
PT acid residues in polypeptide(s)
XX
XX PS Disclosure: Page 13-15; 25pp; Japanese.
XX
XX The sequences given in AA027987-88 encode proteases which were isolated
CC from Staphylococcus aureus strains. The DNA sequences were isolated
CC by PCR using the primer sequences given in AA027960-86. The protease
CC specifically cleaves the peptide bond at the C-terminus of the
CC glutamic acid residue in polypeptide.
XX
XX SQ Sequence 1586 BP; 590 A; 302 C; 219 G; 475 T; 0 other;

Alignment Scores:
Pred. No.: 34.4 Length: 1586
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-10-008-355-26 (1-9) x AA027987 (1-1586)
QY 1 ThrGlyGlyAsnSerGlySerProval 9
DB 1048 ACTGTGTGTAAGTCAAGTTCACCACTA 1074

RESULT 4
ID AAL43635
AC AAL43635 standard; DNA: 2139 BP.
XX
XX
XX
XX 05-SEP-2002 (first entry)
XX
XX DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.
XX
XX DE Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;
XX KM DPP-7 inhibitor identification; periodontal disease; gingivitis;
XX KM periodontitis.
XX
XX OS Porphyromonas gingivalis.
XX
XX FH Key Location/Qualifiers
XX FT 1..2139
XX FT CDS /*tag= a
XX FT /product= "Porphyromonas gingivalis DPP-7"
XX
XX PN MO200238742-A2.
XX
XX PD 16-MAY-2002.
XX
XX PF 08-NOV-2001; 2001MO-US46782.
XX
XX PR 08-NOV-2000; 2000US-246827P.
XX
XX PA (UYGE-) UNIV GEORGIA RES FOUND INC.

XX
XX Travis J, Potempa JS, Banbula A, Bugno M;
XX
XX WPI: 2002-490075/52.
XX P-PSDB; AA015205.
XX
XX PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT the dipeptidylpeptidase for protecting an animal from periodontal
PT disease caused by Porphyromonas gingivalis -
XX
XX PS Claim 11; Fig 4; 65pp; English.
XX
XX The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present DNA sequence encodes the Porphyromonas
CC gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the invention.
XX
XX SQ Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other;

Alignment Scores:
Pred. No.: 47.3 Length: 2139
Score: 48.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 24 Gaps: 0

US-10-008-355-26 (1-9) x AAL43635 (1-2139)
QY 1 ThrGlyGlyAsnSerGlySerProval 9
DB 1930 ACGGGCGGTAACTCGGTAGCCCGCTA 1956

RESULT 5
ID AAH52485
AC AAH52485 standard; DNA: 849 BP.
XX
XX
XX
XX 03-SEP-2001 (first entry)
XX
XX DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:363.
XX
XX KM Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX KM vaccination; endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX
XX PN MO200134809-A2.
XX
XX PD 17-MAY-2001.
XX
XX PF 09-NOV-2000; 2000MO-US30782.
XX
XX PR 09-NOV-1999; 99US-0164258.
XX
XX PA (GLAX) GLAXO GROUP LTD.
XX
XX PI Kimerly MJ;
XX
XX DR WPI: 2001-316495/33.
XX P-PSDB; AAG81635.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -

XX Claim 8; Page 138-139; 2188bp; English.
PS
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II) given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 849 BP; 325 A; 122 C; 147 G; 255 T; 0 other;

Alignment Scores:
Pred. No.: 137 Length: 849
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 89.58% Indels: 0
DB: 22 Gaps: 0

US-10-008-355-26 (1-9) x AAH52485 (1-849)

OY 2 GLYGLYASNSerGlySerProval 9
ID 694 GGTTGGAACCTGTGATCTCCAGTA 717
ABN91429
RESULT 6
ID ABN91429 standard; DNA: 936 BP.
XX
AC ABN91429;
XX
DT 24-JUL-2002 (first entry)
XX
DE *Staphylococcus epidermidis* ORF nucleic acid sequence SEQ ID NO:892.
XX
XX *Staphylococcus epidermidis*; open reading frame; ORF; bacterial infection;
KM antibacterial; gene therapy; gene; ds.
XX
OS *Staphylococcus epidermidis*.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-055779P.
XX
PR 08-NOV-1997; 97US-064964P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
XX
DR WPI: 2002-381255/41.
XX
DR P-Psdb; ABP38884.
XX
PT Novel isolated nucleic acid encoding a *Staphylococcus epidermidis*
XX polypeptide, useful for diagnosing and treating bacterial infections -

WPI
2002-381255/41

XX Disclosure; SEQ ID 892; 267bp; English.
PS
XX ABN90538 to ABN93374 represent *Staphylococcus epidermidis* open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 936 BP; 359 A; 128 C; 154 G; 295 T; 0 other;

Alignment Scores:
Pred. No.: 152 Length: 936
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 89.58% Indels: 0
DB: 24 Gaps: 0

US-10-008-355-26 (1-9) x ABN91429 (1-936)

OY 2 GLYGLYASNSerGlySerProval 9
ID 781 GGTTGGAACCTGTGATCTCCAGTA 804
AAH54330/C
AAH54330 standard; DNA: 3189 BP.
XX
AC AAH54330;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3694.
XX
XX *Staphylococcus epidermidis* SRI strain; infection; diagnosis;
KM vaccination; endocarditis; ds.
XX
OS *Staphylococcus epidermidis*.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
XX (GLAXO) GLAXO GROUP LTD.
XX
PI Kimmery WJ;
XX
DR WPI: 2001-316495/33.
XX
PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
XX useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 1307-1308; 2188bp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II) given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.

CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 3189 BP; 1081 A; 596 C; 502 G; 1010 T; 0 other;

Alignment Scores:
Pred. No.: 558 Length: 3189
Score: 43.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 89.58% Indels: 0
DB: Gaps: 0

US-10-008-355-26 (1-9) x AAH54330 (1-3189)

OY 2 gLyGlyAsnSerGlySerProVal 9
ID 1147 GGTGGAACTCTGCATCTCCAGTA 1124
DB 1147 GGTGGAACTCTGCATCTCCAGTA 1124

RESULT 8
AAAX20213 standard; DNA; 758 BP.
XX
AC AAX20213;
XX
DT 20-APR-1999 (first entry)
XX
XX Enterococcus faecalis EF110 gene fragment.
DE
XX Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KM detection; attenuation; antigenic; ss.
XX
OS Enterococcus faecalis.
XX
PN WO9850554-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98WO-US08959.
XX
PR 14-NOV-1997; 97US-0066009.
XX
PR 06-MAY-1997; 97US-0044031.
XX
PR 16-MAY-1997; 97US-0046655.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
XX
DR MPI: 1999-070095/06.
XX
DR P-PSDB; AAY00223.
XX
PT New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection
XX
XX
PS Claim 1; Page 212-213; 301pp; English.
XX
XX The present sequence encodes an antigenic polypeptide fragment
XX isolated from Enterococcus faecalis. The present invention describes
XX genes, proteins and antigenic polypeptides isolated from *E. faecalis*.
XX The proteins can be used in vaccines for preventing or attenuating an
XX infection caused by a member of the Enterococcus genus in an animal.
XX They can also be used for detecting Enterococcus antibodies in a sample.

CC The nucleotide sequences can be used for detecting Enterococcus nucleic
CC acids. Products from the present invention can also be used for
CC screening compounds to identify agonists and antagonists of *E. faecalis*
CC protein activity.
XX
SQ Sequence 758 BP; 248 A; 143 C; 160 G; 207 T; 0 other;

Alignment Scores:
Pred. No.: 275 Length: 758
Score: 41.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 85.42% Indels: 0
DB: Gaps: 0

US-10-008-355-26 (1-9) x AAX20213 (1-758)

OY 1 ThrGlyAsnSerGlySerProVal 9
ID 575 ACCGGCGTCATCTGTTCCACCAATC 601
DB 575 ACCGGCGTCATCTGTTCCACCAATC 601

RESULT 9
ABN98198 standard; DNA; 758 BP.
XX
AC ABN98198;
XX
DT 05-AUG-2002 (first entry)
XX
XX *E. faecalis* EF110 gene fragment.
DE
XX Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
KM gene; ds.
XX
XX Enterococcus faecalis.
OS
XX
PN US2002045737-A1.
XX
PD 18-APR-2002.
XX
PF 04-MAY-1998; 98US-0071035.
XX
PR 04-MAY-1998; 98US-0071035.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
DR MPI: 2002-425450/45.
XX
DR P-PSDB; ABP43442.
XX
PT New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly *E.*
PT faecalis -
XX
XX
PS Claim 1; Page 197-198; 255pp; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of polypeptides from Enterococcus faecalis. The proteins can be
XX used as vaccines for preventing or attenuating an infection caused by a
XX member of the Enterococcus genus in an animal, particularly *E. faecalis*.
XX The polynucleotide is also useful for preventing or treating *E. faecalis*
XX infection. The present sequence is a coding sequence of the invention.
XX
SQ Sequence 758 BP; 248 A; 143 C; 160 G; 207 T; 0 other;

Alignment Scores:
Pred. No.: 275 Length: 758
Score: 41.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 85.42% Indels: 0

```
DB: 24 Gaps: 0
US-10-008-355-26 (1-9) x ABN98198 (1-758)
QY 1 ThnGlyGlyAsnSerGlySerProval 9
DB 575 ACCGGCGTCATCTGTTCCACCAATC 601

RESULT 10
AAx20212
ID AAx20212 standard; DNA; 888 BP.
XX
AC AAx20212;
XX
DT 20-APR-1999 (first entry)
XX
DE Enterococcus faecalis gene EF110.
XX
KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
KM detection; attenuation; antigenic; ss.
XX
OS Enterococcus faecalis.
XX
PN WO9850554-A2.
XX
PD 12-NOV-1998.
XX
PF 04-MAY-1998; 98MO-US08959.
XX
PR 14-NOV-1997; 97US-0066009.
PR 06-MAY-1997; 97US-0044031.
PR 16-MAY-1997; 97US-0046555.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
XX
DR WPI: 1999-070095/06.
DR P-PSDB; AAy00222.
XX
PT New isolated Enterococcus faecalis polynucleotides - used to develop
PT products for the detection of Enterococcus and for use in vaccines
PT for prevention or attenuation of Enterococcus infection
XX
PS Claim 1; Page 212; 301pp; English.
XX
CC The present sequence represents a gene isolated from
CC Enterococcus faecalis. The present invention describes genes, proteins
CC and antigenic polypeptides isolated from E. faecalis. The proteins can
CC be used in vaccines for preventing or attenuating an infection caused
CC by a member of the Enterococcus genus in an animal. They can also be
CC used for detecting Enterococcus antibodies in a sample. The nucleotide
CC sequences can be used for detecting Enterococcus nucleic acids.
CC Products from the present invention can also be used for screening
CC compounds to identify agonists and antagonists of E. faecalis protein
CC activity.
XX
SQ Sequence 888 BP; 290 A; 159 C; 186 G; 253 T; 0 other;

Alignment Scores:
Pred. No.: 326 Length: 888
Score: 41.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 85.42% Indels: 0
DB: 20 Gaps: 0

US-10-008-355-26 (1-9) x AAx20212 (1-888)
QY 1 ThnGlyGlyAsnSerGlySerProval 9
DB 703 ACCGGCGTCATCTGTTCCACCAATC 729
```

```
RESULT 11
ABN98197
ID ABN98197 standard; DNA; 888 BP.
XX
AC ABN98197;
XX
DT 05-AUG-2002 (first entry)
XX
DE E faecalis EF110 gene.
XX
KW Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic;
KM gene; ds.
XX
OS Enterococcus faecalis.
XX
PN US2002045737-A1.
XX
PD 18-APR-2002.
XX
PF 04-MAY-1998; 98US-0071035.
XX
PR 04-MAY-1998; 98US-0071035.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
DR WPI: 2002-425450/45.
DR P-PSDB; ABP43441.
XX
PT New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly E.
PT faecalis
XX
PS Claim 1; Page 197; 255pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of polypeptides from Enterococcus faecalis. The proteins can be
CC used as vaccines for preventing or attenuating an infection caused by a
CC member of the Enterococcus genus in an animal, particularly E. faecalis.
CC The polynucleotide is also useful for preventing or treating E. faecalis
CC infection. The present sequence is a coding sequence of the invention.
XX
SQ Sequence 888 BP; 290 A; 159 C; 186 G; 253 T; 0 other;

Alignment Scores:
Pred. No.: 326 Length: 888
Score: 41.00 Matches: 7
Percent Similarity: 88.89% Conservative: 1
Best Local Similarity: 77.78% Mismatches: 1
Query Match: 85.42% Indels: 0
DB: 24 Gaps: 0

US-10-008-355-26 (1-9) x ABN98197 (1-888)
QY 1 ThnGlyGlyAsnSerGlySerProval 9
DB 703 ACCGGCGTCATCTGTTCCACCAATC 729

RESULT 12
ABQ91212/C
ID ABQ91212 standard; DNA; 2130 BP.
XX
AC ABQ91212;
XX
DT 01-OCT-2002 (first entry)
XX
DE M. capsulatus gene #1197 for DNA array.
XX
KW Micro array; gene; ds; differential expression; gene expression.
XX
OS Methylococcus capsulatus.
```



```
XX XX WO200255655-A2.
XX PN
XX 18-JUL-2002.
XX PD
XX PF 14-JAN-2002; 2002WO-NO00019.
XX PR 12-JAN-2001; 2001NO-0000235.
XX PR 12-JAN-2001; 2001NO-0000239.
XX PA (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.
XX PA (TIGR-) TIGR.
XX PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;
XX PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;
XX PI Salzberg SL;
XX DR WPI; 2002-557818/59.
XX PT Novel DNA array useful for determining differential expression of
XX PT Methyloloboccus capsulatus genes, comprises polynucleotides or
XX PT oligonucleotides representative for a selective number of Methyloloboccus
XX PT capsulatus genes
XX PS Claim 14; Page 497; 678pp; English.
XX CC The invention relates to a novel DNA array giving a representation of a
XX CC number of Methyloloboccus capsulatus genes. The method of the invention is
XX CC useful for determination of the differential expression of the genes of
XX CC M. capsulatus, and for studying gene expression on a genomic scale and in
XX CC gene expression assays of M. capsulatus genes. The sequences shown in
XX CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the
XX CC invention.
XX SQ Sequence 2130 BP; 363 A; 633 C; 705 G; 429 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 824 Length: 2130
XX Score: 41.00 Matches: 7
XX Percent Similarity: 100.00% Conservative: 1
XX Best Local Similarity: 87.50% Mismatches: 0
XX Query Match: 85.42% Indels: 0
XX DB: 24 Gaps: 0
XX
XX US-10-008-355-26 (1-9) x ABQ91212 (1-2130)
XX
XX QY 1 ThrGlyGlyAsnSerGlySerPro 8
XX ID |||||||||:::|||||||
XX DB 1661 ACCGGCGGCAACGCTGCTGCCG 1638
XX
XX RESULT 13
XX AAX13198/C
XX ID AAX13198 standard; DNA; 4951 BP.
XX AC AAX13198;
XX DT 19-MAR-1999 (first entry)
XX DE Enterococcus faecalis genome contig SRQ ID NO:261.
XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;
XX KW vaccine; attenuation; computer readable medium; ds.
XX OS Enterococcus faecalis.
XX PN WO9850555-A2.
XX PD 12-NOV-1998.
XX PF 04-MAY-1998; 98WO-US08985.
XX PR 14-NOV-1997; 97US-0066009.
XX PR 06-MAY-1997; 97US-0044031.
```

```
PR 16-MAY-1997; 97US-0046655.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Barash SC, Dillon PJ, Kunsch CA;
XX DR WPI; 1999-045171/04.
XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides
XX PT - used to develop products for the detection of Enterococcus
XX PT use in vaccines for prevention or attenuation of Enterococcus
XX PT infection.
XX PS Claim 1; Page 1248-1251; 2084pp; English.
XX CC A computer readable medium has been developed which has recorded on it
XX CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
XX CC AAX12938 to AAX13919 represent these nucleotide sequences which are
XX CC primary nucleotide sequences, also known as contigs. The computer-based
XX CC system can identify fragments of the Enterococcus faecalis genome with
XX CC commercial importance. The products can be used to detect the presence
XX CC of Enterococcus faecalis in samples. They can also be used for
XX CC diagnosing Enterococcal infection in an animal and monitoring
XX CC progression of disease, and for identifying agents which can be used to
XX CC modulate the growth or pathogenicity of Enterococcus faecalis, or
XX CC another related organism, in vivo or in vitro. In particular the
XX CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
XX CC can be used in vaccines to prevent or attenuate an Enterococcal
XX CC infection.
XX SQ Sequence 4951 BP; 1563 A; 981 C; 735 G; 1665 T; 7 other;
XX
XX Alignment Scores:
XX Pred. No.: 2,02e+03 Length: 4951
XX Score: 41.00 Matches: 7
XX Percent Similarity: 88.89% Conservative: 1
XX Best Local Similarity: 77.78% Mismatches: 1
XX Query Match: 85.42% Indels: 0
XX DB: 20 Gaps: 0
XX
XX US-10-008-355-26 (1-9) x AAX13198 (1-4951)
XX
XX QY 1 ThrGlyGlyAsnSerGlySerProval 9
XX ID ||||||||| |||||||||
XX DB 695 ACCGGCGGCTCAATCTGTTCCACCAATC 669
XX
XX RESULT 14
XX ABL78727
XX ID ABL78727 standard; cDNA; 450 BP.
XX AC ABL78727;
XX DT 17-MAY-2002 (first entry)
XX DE Human ovarian cancer related cDNA clone SEQ ID NO:1705.
XX KW Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
XX OS Homo sapiens.
XX PN WO200192581-A2.
XX PD 06-DEC-2001.
XX PF 29-MAY-2001; 2001WO-US17756.
XX PR 26-MAY-2000; 2000US-207484P.
XX PA (CORI-) CORIXA CORP.
XX PI Algate PA, Harlocker SL, Jones R;
XX DR WPI; 2002-122075/16.
```

XX Composition for therapy and diagnosis of ovarian cancer comprising
 PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 PT polypeptide, antibody specific to polypeptide or T cell expressing
 PT polypeptide
 XX
 PS Claim 1, SEQ ID 1705; 489pp; English.
 XX
 CC The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumor
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence
 CC (S1) from the 10912 nucleotide sequences as given in ABL77023 to
 CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell
 CC population of (II), or antigen presenting cells that express (II).
 CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to
 CC (S1) can be used for detecting ovarian cancer in a patient's biological
 CC sample preferably serum or ovarian tissue. The method comprises
 CC contacting a biological sample from a patient with (IV), detecting the
 CC amount of polynucleotide hybridising to (IV) and comparing the amount to
 CC a predetermined cutoff value and thereby detecting ovarian cancer in the
 CC patient, where the amount of polynucleotide hybridising to (IV) is
 CC detected preferably by polymerase chain reaction (PCR). (I) comprising
 CC (III) and/or (II) is useful for stimulating and/or expanding T cells
 CC specific for an ovarian tumor protein comprising contacting T cells
 CC with (III) or (II). (III) is useful in design and preparation of
 CC ribozyme molecules for inhibiting expression of the tumor polypeptides
 CC and proteins in tumor cells; and to isolate a full length gene from a
 CC suitable library e.g., a tumor cDNA library using well known
 CC techniques.
 CC
 SQ Sequence 450 BP; 148 A; 109 C; 109 G; 81 T; 3 other;
 XX
 XX
 Alignment Scores:
 Pred. No.: 238 Length: 450
 Score: 40.00 Matches: 7
 Percent Similarity: 87.50% Conservative: 0
 Best Local Similarity: 87.50% Mismatches: 1
 Query Match: 83.33% Indels: 0
 DB: 24 Gaps: 0
 US-10-008-355-26 (1-9) x ABL78727 (1-450)
 OY 1 ThrGlyGlyAsnSerGlySerPro 8
 |||||
 DB 82 ACAGAGAGTAATCTGCGCAAGCC 105
 RESULT 15
 AAF07536/C
 ID AAF07536 standard; cDNA; 1127 BP.
 XX
 AC AAF07536;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:59.
 XX
 KM Multiple gene expression: filamentous fungal cell; EST:
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; Identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000MO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berta RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX
 DR WPI: 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 393-394; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 1127 BP; 232 A; 356 C; 268 G; 258 T; 13 other;
 XX
 XX
 Alignment Scores:
 Pred. No.: 632 Length: 1127
 Score: 40.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 87.50% Mismatches: 0
 Query Match: 83.33% Indels: 0
 DB: 21 Gaps: 0
 US-10-008-355-26 (1-9) x AAF07536 (1-1127)
 OY 2 GlyGlyAsnSerGlySerProVal 9
 |||||
 DB 332 GGAGGGAAGTCAAGAACACCAAGTG 309

Search completed: December 20, 2002, 13:22:54
 Job time : 175.474 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 09:34:23 : Search time 15.6316 Seconds
(Without alignments)
16.940 Million cell updates/sec

Title: US-10-008-355-26

Perfect score: 48

Sequence: 1 TGNNSGSPV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	15	2	US-08-332-562A-67
2	48	100.0	213	3	US-08-523-373-22
3	48	100.0	214	3	US-08-523-373-23
4	48	100.0	215	3	US-08-523-373-24
5	48	100.0	344	1	US-08-657-192-3
6	48	100.0	344	3	US-08-523-373-5
7	48	100.0	392	3	US-08-523-373-6
8	48	100.0	532	1	US-08-657-192-9
9	48	100.0	532	3	US-08-523-373-7
10	48	100.0	537	1	US-08-657-192-15
11	43	89.6	311	4	US-09-134-001C-3729
12	41	85.4	252	4	US-09-071-035-428
13	41	85.4	284	4	US-09-071-035-426
14	39	81.2	418	4	US-09-342-653-7
15	37	77.1	222	1	US-08-090-048-1
16	37	77.1	222	2	US-08-292-550-1
17	37	77.1	222	2	US-07-927-661A-1
18	36	75.0	379	1	US-07-723-002C-6
19	36	75.0	393	4	US-09-194-905-13
20	36	75.0	628	2	US-07-952-853-22
21	35	72.9	628	2	US-08-914-848-22
22	35	72.9	318	4	US-09-060-756-727
23	35	72.9	334	4	US-09-060-756-728
24	35	72.9	551	2	US-09-033-537A-1
25	34	70.8	376	4	US-09-056-556-202
26	34	70.8	376	4	US-09-072-596-197
27	34	70.8	501	2	US-08-660-963-13

28	34	70.8	501	2	US-08-969-630-4	Sequence 4, Appl
29	34	70.8	502	2	US-08-969-630-2	Sequence 2, Appl
30	33	68.8	44	1	US-08-450-945-69	Sequence 69, Appl
31	33	68.8	44	4	US-08-976-161-69	Sequence 69, Appl
32	33	68.8	228	4	US-09-219-849-38	Sequence 38, Appl
33	33	68.8	379	2	US-08-887-365-36	Sequence 36, Appl
34	33	68.8	385	1	US-08-361-920-23	Sequence 23, Appl
35	33	68.8	385	1	US-08-479-939-23	Sequence 23, Appl
36	33	68.8	385	1	US-08-483-432-23	Sequence 23, Appl
37	32	66.7	9	1	US-08-638-911A-8	Sequence 8, Appl
38	32	66.7	43	3	US-08-946-026-5	Sequence 5, Appl
39	32	66.7	63	4	US-08-817-787-17	Sequence 17, Appl
40	32	66.7	108	1	US-08-466-033-106	Sequence 106, App
41	32	66.7	108	2	US-08-444-733-106	Sequence 106, App
42	32	66.7	108	2	US-08-464-134-106	Sequence 106, App
43	32	66.7	108	2	US-08-461-361-106	Sequence 106, App
44	32	66.7	108	2	US-08-485-910-106	Sequence 106, App
45	32	66.7	108	5	PCT-US95-06266-89	Sequence 89, Appl

ALIGNMENTS

RESULT 1
US-08-332-562A-67
Sequence 67, Application US/08332562A
Patent No. 5985599
GENERAL INFORMATION:
APPLICANT: MCKENZIE, Ian F.C.
APPLICANT: HOGARTH, Mark P.
APPLICANT: HIBBS, Margaret L.
APPLICANT: SCOTT, Bernadette M.
APPLICANT: BONADONNA, Lisa
APPLICANT: HOLETT, Mark D.
TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN
NUMBER OF SEQUENCES: 136
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,562A
FILING DATE: 31-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/896,457
FILING DATE: 27-MAY-1992
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 54270/119/GRHA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-332-562A-67
Query Match 100.0%; Score 48; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.083;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGGNSGSPV 9
 |||||
 Db 5 TGGNSGSPV 13

16

RESULT 2
 US-08-523-373-22
 : Sequence (22) Application US/08523373
 : Patent NO. 6037145
 : GENERAL INFORMATION:
 : APPLICANT: Yabuta, Masayuki
 : APPLICANT: Ohnuye, Kazuhiro
 : TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
 : NUMBER OF SEQUENCES: 24
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
 : STREET: 699 Prince Street
 : CITY: Alexandria
 : STATE: VA
 : COUNTRY: USA
 : ZIP: 22314-3187
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/523,373
 : FILING DATE: 05-SEP-1995
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 6-238595
 : FILING DATE: 07-SEP-1994
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 6-296028
 : FILING DATE: 07-NOV-1994
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Neuth, Donna M.
 : REGISTRATION NUMBER: 36,607
 : REFERENCE/DOCKET NUMBER: 001560-251
 : TELEPHONE: 703-836-6620
 : TELEFAX: 703-836-2021
 : INFORMATION FOR SEQ ID NO: 22:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 213 amino acids
 : TYPE: amino acid
 : STRANDEDNESS: not relevant
 : TOPOLOGY: linear
 : MOLECULE TYPE: peptide
 : US-08-523-373-22

Query Match 100.0%; Score 48; DB 3; Length 213;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Same pub as #2

RESULT 3
 US-08-523-373-23
 : Sequence (23) Application US/08523373
 : Patent NO. 6037145
 : GENERAL INFORMATION:
 : APPLICANT: Yabuta, Masayuki
 : APPLICANT: Ohnuye, Kazuhiro
 : TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
 : NUMBER OF SEQUENCES: 24
 : CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
 STREET: 699 Prince Street
 CITY: Alexandria
 STATE: VA
 COUNTRY: USA
 ZIP: 22314-3187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/523,373
 FILING DATE: 05-SEP-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 6-238595
 FILING DATE: 07-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: JP 6-296028
 FILING DATE: 07-NOV-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Neuth, Donna M.
 REGISTRATION NUMBER: 36,607
 REFERENCE/DOCKET NUMBER: 001560-251
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-836-6620
 TELEFAX: 703-836-2021
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 214 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-523-373-23

#6,7

Query Match 100.0%; Score 48; DB 3; Length 214;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4
 US-08-523-373-24
 : Sequence (24) Application US/08523373
 : Patent NO. 6037145
 : GENERAL INFORMATION:
 : APPLICANT: Yabuta, Masayuki
 : APPLICANT: Ohnuye, Kazuhiro
 : TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
 : NUMBER OF SEQUENCES: 24
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
 : STREET: 699 Prince Street
 : CITY: Alexandria
 : STATE: VA
 : COUNTRY: USA
 : ZIP: 22314-3187
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/523,373
 : FILING DATE: 05-SEP-1995
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: JP 6-238595

FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028
FILING DATE: 07-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 215 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-24

Query Match 100.0%; Score 48; DB 3; Length 215;
Best Local Similarity 100.0%; Pred. NO. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||
DB 165 TCGNSGSPV 173

RESULT 5
US-08-657-192-3
Sequence 3, Application US/08657192

Patent No. 5747321

GENERAL INFORMATION:
APPLICANT: TABUTA, Masayuki
APPLICANT: OHSUYE, Kazuhiro
TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
TITLE OF INVENTION: PROTEASES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/657,192
FILING DATE: 03-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-170086
FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-657-192-3

Query Match 100.0%; Score 48; DB 1; Length 344;
Best Local Similarity 100.0%; Pred. NO. 1.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||
DB 289 TCGNSGSPV 297

RESULT 6
US-08-523-373-5
Sequence 5, Application US/08523373
Patent No. 6037145

GENERAL INFORMATION:
APPLICANT: Tabuta, Masayuki
APPLICANT: Ohsuye, Kazuhiro
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
STREET: 699 Prince Street
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22314-3187

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/523,373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-238595
FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028

FILING DATE: 07-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 344 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-5

Query Match 100.0%; Score 48; DB 3; Length 344;
Best Local Similarity 100.0%; Pred. NO. 1.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||
DB 289 TCGNSGSPV 297

RESULT 7
US-08-523-373-6

Sequence 6, Application US/08523373
Patent No. 6037145
GENERAL INFORMATION:
APPLICANT: Yabuta, Masayuki
APPLICANT: Ohsuye, Kazuhiro

;; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
;; NUMBER OF SEQUENCES: 24
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
;; STREET: 699 Prince Street
;; CITY: Alexandria
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22314-3187
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/523,373
;; FILING DATE: 05-SEP-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 6-238595
;; FILING DATE: 07-SEP-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 6-296028
;; FILING DATE: 07-NOV-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meuth, Donna M.
;; REGISTRATION NUMBER: 36,607
;; REFERENCE/DOCKET NUMBER: 001560-251
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-836-6620
;; TELEFAX: 703-836-2021
;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 392 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-523-373-6

Query Match 100.0%; Score 48; DB 3; Length 392;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9
DB 289 TGGNSGSPV 297

RESULT 8
US-08-657-192-9
;; Sequence 9, Application US/08657192
;; Patent No. 5747321
;; GENERAL INFORMATION:
;; APPLICANT: YABUTA, Masayuki
;; APPLICANT: OHSUYE, Kazuhiko
;; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
;; TITLE OF INVENTION: PROTEASES
;; NUMBER OF SEQUENCES: 15
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
;; STREET: P.O. Box 1404
;; CITY: Alexandria
;; STATE: Virginia
;; COUNTRY: United States
;; ZIP: 22313-1404
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/657,192

;; FILING DATE: 03-JUN-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 7-170086
;; FILING DATE: 02-JUN-1995
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meuth, Donna M.
;; REGISTRATION NUMBER: 36,607
;; REFERENCE/DOCKET NUMBER: 001560-264
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 836-6620
;; TELEFAX: (703) 836-2021
;; INFORMATION FOR SEQ ID NO: 9:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 532 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-657-192-9

Query Match 100.0%; Score 48; DB 1; Length 532;
Best Local Similarity 100.0%; Pred. No. 2,7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9
DB 289 TGGNSGSPV 297

RESULT 9
US-08-523-373-7
;; Sequence 7, Application US/08523373
;; Patent No. 6037145
;; GENERAL INFORMATION:
;; APPLICANT: Yabuta, Masayuki
;; APPLICANT: Ohsuye, Kazuhiko
;; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF PROTEIN
;; NUMBER OF SEQUENCES: 24
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
;; STREET: 699 Prince Street
;; CITY: Alexandria
;; STATE: VA
;; COUNTRY: USA
;; ZIP: 22314-3187
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/523,373
;; FILING DATE: 05-SEP-1995
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 6-238595
;; FILING DATE: 07-SEP-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: JP 6-296028
;; FILING DATE: 07-NOV-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Meuth, Donna M.
;; REGISTRATION NUMBER: 36,607
;; REFERENCE/DOCKET NUMBER: 001560-251
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 703-836-6620
;; TELEFAX: 703-836-2021
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 532 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-523-373-7
Query Match
Best Local Similarity 100.0%; Score 48; DB 3; Length 532;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 9
Db 289 TGNSSGSPV 297

RESULT 10
US-08-657-192-15
; Sequence 15, Application US/08657192
; Patent No. 5747321
; GENERAL INFORMATION:
; APPLICANT: YABUTA, Masayuki
; APPLICANT: OHSUKE, Kazuhito
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/657,192
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-170086
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Neuth, Donna M.
; REGISTRATION NUMBER: 36,607
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-657-192-15
Query Match
Best Local Similarity 100.0%; Score 48; DB 1; Length 537;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 9
Db 289 TGNSSGSPV 297

RESULT 11
US-09-134-001C-3729
; Sequence 3729, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
```

```

;
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3729
; LENGTH: 311
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3729
Query Match
Best Local Similarity 89.6%; Score 43; DB 4; Length 311;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGNSSGSPV 9
Db 261 GGNSSGSPV 268

RESULT 12
US-09-071-035-428
; Sequence 428, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: GILL H. CHOI
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 428:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 252 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-428
Query Match
Best Local Similarity 85.4%; Score 41; DB 4; Length 252;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 9
```

Db 192 TCGGSGSPI 200

RESULT 13

US-09-071-035-426
; Sequence 426, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071.035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 426:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 284 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-426

Query Match 85.4%; Score 41; DB 4; Length 284;
Best Local Similarity 77.8%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
Db 224 TCGGSGSPI 232

RESULT 14

US-09-342-653-7
; Sequence 7, Application US/09342653
; Patent No. 6306632
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Chromatin Associated Proteins
; FILE REFERENCE: BB-1118
; CURRENT APPLICATION NUMBER: US/09/342,653
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: 60/092,841
; EARLIER FILING DATE: July 14, 1998
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 7
; LENGTH: 418

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-342-653-7

Query Match 81.2%; Score 39; DB 4; Length 418;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGNCGSPV 9
Db 163 GGNCGSPI 170

RESULT 15

US-08-090-048-1
; Sequence 1, Application US/08090048
; Patent No. 5523237
; GENERAL INFORMATION:
; APPLICANT: Budtz, Peter
; APPLICANT: Nielsen, Per M.
; TITLE OF INVENTION: PROTEIN PREPARATIONS
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 55232370 No. 5523237disk of No. 5523237th America, Inc.
; STREET: 405 Lexington Avenue, 62nd Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/090,048
; FILING DATE: 16-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 199/91
; FILING DATE: 06-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK92/00036
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H.
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 3396.214-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-867-0298
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-090-048-1

Query Match 77.1%; Score 37; DB 1; Length 222;
Best Local Similarity 87.5%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGNCGSPV 9
Db 164 GGNCGSPV 171

Search completed: December 20, 2002, 12:12:19
Job time : 16.6316 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 12:12:24 ; Search time 1368.95 seconds

(without alignments)
106.476 Million cell updates/sec

Title: US-10-008-355-26

Perfect score: 48

Sequence: 1 TCGNSGSPV 9

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_plus.p2n.model -DEV=x1h
-O=/cgn2.1/USPTO.spool/US10008355/runat.17122002.112338.14609/app.query.fasta.1.398
-DB=EST -QFMT=fastap -SUFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355 @cgn2.1.1.763 @runat.17122002.112338.14609 -NCPD=6 -ICPD=3
-NO_XLPXY -NO_MAP -LARGEBOOKER -NEG.SCORES=0 -WAIT -DONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estlmu: *
5: em_estlov: *
6: em_estlpl: *
7: em_estro: *
8: em_est: *
9: gb_estl: *
10: gb_estl2: *
11: gb_hcc: *
12: gb_estl3: *
13: gb_estl4: *
14: gb_estl5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pin: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rtd: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	47	97.9	318	9	AA411980	AA411980 zt65g03.s
2	47	97.9	426	17	BH400391	BH400391 AG-ND-147
3	45	93.8	371	17	AO699991	AO699991 HS_5334_A
4	45	93.8	796	12	BC714149	BC714149 602674594
5	44	91.7	429	10	AM060086	AM060086 687001007
6	44	91.7	420	14	W72307	W72307 z660g08.s1
7	44	91.7	768	13	BC918747	BC918747 602819216
8	44	91.7	958	14	BO881134	BO881134 AGENCOURT
9	44	91.7	970	12	BE794754	BE794754 601590584
10	43	89.6	1466	13	BI545222	BI545222 603187343
11	42	87.5	185	9	AI608160	AI608160 val14g09.y
12	42	87.5	346	9	AA242434	AA242434 MX26c10.1
13	42	87.5	405	9	AI451721	AI451721 MX26c10.x
14	42	87.5	426	9	AI661302	AI661302 val14g09.x
15	42	87.5	456	9	AA964075	AA964075 UI-R-CO-9
16	42	87.5	473	12	BF397465	BF397465 UI-R-BB2-
17	42	87.5	475	12	BF390358	BF390358 UI-R-CAI-
18	42	87.5	484	12	BF467444	BF467444 UI-M-CGP
19	42	87.5	495	9	AI029930	AI029930 UI-R-CO-1
20	42	87.5	576	17	BH300672	BH300672 CH230-200
21	42	87.5	714	10	BM383120	BM383120 BM383120
22	42	87.5	983	13	BM465277	BM465277 AGENCOURT
23	41	85.4	225	10	AW812682	AW812682 RCO-ST018
24	41	85.4	469	9	AL795494	AL795494 AL795494
25	41	85.4	608	10	BB622700	BB622700 BB622700
26	41	85.4	619	9	AL785210	AL785210 AL785210
27	41	85.4	651	14	BQ397257	BQ397257 NISC-ng26
28	41	85.4	680	9	AL647551	AL647551 AL647551
29	41	85.4	682	17	AG083764	AG083764 Pan trogl
30	41	85.4	692	9	AL654837	AL654837 AL654837
31	41	85.4	704	9	AL647370	AL647370 AL647370
32	41	85.4	764	10	AW174380	AW174380 f142604.y
33	41	85.4	787	13	BI737468	BI737468 603358149
34	41	85.4	894	17	CNS03JF1	AL246790 Tetradon
35	41	85.4	914	12	BF128636	BF128636 601810895
36	41	85.4	995	12	BE882829	BE882829 601509232
37	41	85.4	1089	10	BE512670	BE512670 601171626
38	40	83.3	225	10	BB589647	BB589647 BB589647
39	40	83.3	377	10	AW308476	AW308476 3608 MARC
40	40	83.3	388	10	AW148054	AW148054 da03d08.x
41	40	83.3	391	9	AA413994	AA413994 vc64d04.s
42	40	83.3	392	10	AW315936	AW315936 13723 MAR
43	40	83.3	399	10	AW147471	AW147471 da03d08.y
44	40	83.3	414	12	BG834820	BG834820 353172 MA
45	40	83.3	418	13	BI187317	BI187317 a2d04fs.r

ALIGNMENTS

RESULT 1
AA411980/c 318 bp mRNA linear EST 12-AUG-1997
LOCUS zt65g03.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:727252
DEFINITION
AA411980
VERSION AA411980.1 GI:2070632
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 318)
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisels, G., Jost, S.,
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.

TITLE
JOURNAL
COMMENT
Schellenberg, K., Steptoe, M., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2072 Std Error: 0.00
Seq primer: -41ml3 fwd, ET from Amersham
High quality sequence stop: 306.

FEATURES

source

1..318
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:592416"
/db_xref="taxon:9606"
/clone="IMAGE:727252"
/clone.lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5'] TGTTCACATCTGACGTGGAGCGGCCGCACATTTTCTTTTCTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

90 a 72 c 58 g 98 t

ALIGNMENT SCORES:

Pred. No.: 268 Length: 318
Score: 47.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 97.92% Indels: 0
DB: 9 Gaps: 0

US-10-008-355-26 (1-9) x AA411980 (1-318)

QY 1 ThrcGlyGlyAsnSerGlySerProval 9

DB 219 ACGGAGGAGAACTGTGCACTCCTATT 193

RESULT 2

BH400391/c 426 bp DNA linear GSS 11-DIC-2001

LOCUS AG-ND-147H4.TF ND-TAM Anopheles gambiae genomic clone AG-ND-147H4,

DEFINITION DNA sequence.

ACCESSION BH400391

VERSION BH400391.1 GI:17346607

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

REFERENCE 1 (bases 1 to 426)

AUTHORS Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL Unpublished (2001)

COMMENT Other GSSs: AG-ND-147H4.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@slgr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 For
Class: BAC ends.

FEATURES

source

1..426
Location/Qualifiers
/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="AG-ND-147H4"
/clone.lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"

BASE COUNT

111 a 79 c 76 g 160 t

ALIGNMENT SCORES:

Pred. No.: 356 Length: 426
Score: 47.00 Matches: 8
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 88.89% Mismatches: 0
Query Match: 97.92% Indels: 0
DB: 17 Gaps: 0

US-10-008-355-26 (1-9) x BH400391 (1-426)

QY 1 ThrcGlyGlyAsnSerGlySerProval 9

DB 81 ACAGAGGTACTACAGTCTCTCCATT 55

RESULT 3

AO699991 371 bp DNA linear GSS 06-JUL-1999

LOCUS HS-5334_A2-G04-SP6E RPC1-11 Human Male BAC Library Homo sapiens

DEFINITION genomic clone Plate=910 Col=8 Row=M, DNA sequence.

ACCESSION AO699991

VERSION AO699991.1 GI:5390239

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 371)

AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T., Kellar, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and Hood, L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPC1-11. For BAC library availability, please contact Pictet de Jong

(pictet@dejong.med.buflalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buflalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>

Plate: 910 row: M column: 8

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 10 Gaps: 0

US-10-008-355-26 (1-9) x AW060086 (1-279)

QY 1 ThnglyGlyAsnserGlySerPro 8
|||||
DB 155 ACAGAGGAAACGATGCGACGCCG 132

RESULT 6
W72307 420 bp mRNA linear EST 16-OCT-1996
LOCUS z60g08.s1 Soares_fetal_heart_NbH19W Homo sapiens CDNA clone
DEFINITION IMAGE:345086 3', mRNA sequence.
W72307
KEYWORDS W72307.1 GI:1382930
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 420)
Hiller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Maita, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
JOURNAL
COMMENT Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL: contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 389 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 325.

FEATURES
Location/Qualifiers
1..420
/organism="Homo sapiens"
/db_xref="GDB:1270461"
/db_xref="taxon:9606"
/clone="IMAGE:345086"
/clone_lib="Soares_fetal_heart_NbH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pRT73D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTATGCATCTGAGTGGAGCGCCGCACTCTTTTCTTTTCTTTT 3'], TGTATGCATCTGAGTGGAGCGCCGCACTCTTTTCTTTTCTTTT 3'], RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M. Facina Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbH19W."

BASE COUNT 129 a 121 c 88 g 79 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 1.06e+03 Length: 420
Score: 44.00 Matches: 8
Percent Similarity: 88.89% Conservative: 0
Best Local Similarity: 88.89% Mismatches: 1
Query Match: 91.67% Indels: 0
DB: 14 Gaps: 0

US-10-008-355-26 (1-9) x W72307 (1-420)

QY 1 ThnglyGlyAsnserGlySerProval 9
|||||
DB 201 ACAGGTGGCAACACGCGANTCCCGTG 227

RESULT 7
BG918747 768 bp mRNA linear EST 05-JUN-2001
LOCUS BG918747
DEFINITION 602819216f1 NCI_CGAP_Mam6 Mus musculus CDNA clone IMAGE:4948443 5', mRNA sequence.
BG918747
KEYWORDS BG918747.1 GI:14299223
SOURCE EST.
ORGANISM house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 768)
NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM10899 row: 0 column: 04
High quality sequence stop: 733.

FEATURES
Location/Qualifiers
1..768
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4948443"
/clone_lib="NCI_CGAP_Mam6"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

BASE COUNT 229 a 206 c 203 g 130 t
ORIGIN

Alignment Scores:
Pred. No.: 1.89e+03 Length: 768
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 13 Gaps: 0

US-10-008-355-26 (1-9) x BG918747 (1-768)

QY 1 ThnglyGlyAsnserGlySerPro 8
|||||
DB 422 ACAGAGGTAATTCTGCGACGCCA 445

RESULT 8
BO881134 958 bp mRNA linear EST 16-AUG-2002
LOCUS BO881134
DEFINITION AGENCOURT 8728373 NIH_MGC_47 Homo sapiens CDNA clone IMAGE:6393187 5', mRNA sequence.
BO881134
KEYWORDS BO881134.1 GI:22273142
SOURCE EST.

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1CM2533 row: j column: 20
High quality sequence stop: 504.
Location/Qualifiers

FEATURES
source
1. 958
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6339187"
/clone.lib="NIH_MGC_47"
/tissue.type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC Library."

BASE COUNT 181 a 321 c 293 g 163 t

ORIGIN

Alignment Scores:
Pred. No.: 2.34e+03 Length: 958
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 14 Gaps: 0

US-10-008-355-26 (1-9) x BQ881134 (1-958)

QY 1 ThnGlyGlyAsnSerGlySerPro 8
Db 776 ACCGGCGGAGACTCAGGTCGCCA 799

RESULT 9
LOCUS BE794754 970 bp mRNA linear EST 20-SEP-2000
DEFINITION 601590584F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3944516 5',
mRNA sequence.
ACCESSION BE794754
VERSION BE794754.1 GI:10215952
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTF

CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM801 row: h column: 21
High quality sequence stop: 690.
Location/Qualifiers

FEATURES
source
1. 970
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3944516"
/clone.lib="NIH_MGC_7"
/tissue.type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 181 a 273 c 267 g 248 t 1 others

ORIGIN

Alignment Scores:
Pred. No.: 2.37e+03 Length: 970
Score: 44.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 91.67% Indels: 0
DB: 12 Gaps: 0

US-10-008-355-26 (1-9) x BE794754 (1-970)

QY 1 ThnGlyGlyAsnSerGlySerPro 8
Db 885 ACCGGCGGACACAGTGGTCCCC 862

RESULT 10
LOCUS B1545222 1466 bp mRNA linear EST 05-SEP-2001
DEFINITION 603187333F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5258866 5',
mRNA sequence.
ACCESSION B1545222
VERSION B1545222.1 GI:15432534
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Miklos Palokovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: L1AM1652 row: e column: 11
High quality sequence start: 71
High quality sequence stop: 136.
Location/Qualifiers

FEATURES
source
1. 1466
/organism="Homo sapiens"

```
/db_xref="taxon:9606"
/clone_image="5258866"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dt primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
```

BASE COUNT 394 a 348 c 349 g 375 t

ORIGIN

Alignment Scores:

Pred. No.:	5.1e+03	Length:	1466
Score:	43.00	Matches:	7
Percent Similarity:	88.89%	Conservative:	1
Best Local Similarity:	77.78%	Mismatches:	1
Query Match:	89.58%	Indels:	0
DB:	13	Gaps:	0

US-10-008-355-26 (1-9) x B1545222 (1-1466)

OY 1 ThrGlyGlyAsnSerGlySerProval 9
|||||

Db 1198 ACTGCTGGAATGCTGGAGCTCCTATC 1224 .

RESULT 11

AI608160/c	185 bp	mRNA	linear	EST 21-APR-1999
LOCUS	AA242434			
DEFINITION	val4g09.y1 Soares mouse lymph node NBLMLN Mus musculus cDNA clone			
ACCESSION	AI608160			
VERSION	AI608160.1	GI:4617327		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 185)			
AUTHORS	Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.			
TITLE	The WashU-NCI Mouse EST Project 1999			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watsn.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:444440 This read is a RESSEQUENCE of a previously sequenced mouse clone This read has been verified (found to hit its original self in the Seq primer: -40RP from Gibco High quality sequence stop: 184. Location/Qualifiers 1..185 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone_image="722944" /clone_lib="Soares mouse lymph node NBLMLN" /sex="male"			

```
/tissue_type="lymph node"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pT7n3D-Pac (Pharmacia)  
with a modified polylinker. Site_1: Not I; Site_2: Eco RI;  
1st strand cDNA was primed with a Not I - oligo(dT) primer  
15'  
TGTTCACATCTGAAGTGGAGCGCCGATCACTTTTTTTTTTTTTTTTTTTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7n3 vector. RNA  
provided by Dr. Bertrand Jordan. Library constructed and  
normalized by Bento Soares and M.Fatima Bonaldo."
```

BASE COUNT 37 a 47 c 52 g 49 t

ORIGIN

Alignment Scores:

Pred. No.:	1e+03	Length:	185
Score:	42.00	Matches:	8
Percent Similarity:	88.89%	Conservative:	0
Best Local Similarity:	88.89%	Mismatches:	1
Query Match:	87.50%	Indels:	0
DB:	9	Gaps:	0

US-10-008-355-26 (1-9) x AI608160 (1-185)

OY 1 ThrGlyGlyAsnSerGlySerProval 9
|||||

Db 92 ACAGGAGGAGAGCTCAGGCTCCCTGTC 66

RESULT 12

AA242434	346 bp	mRNA	linear	EST 07-MAR-1997
LOCUS	mx26c10.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:681330			
DEFINITION	5', mRNA sequence.			
ACCESSION	AA242434			
VERSION	AA242434.1	GI:1873137		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 346)			
AUTHORS	Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.			
TITLE	The WashU-HHMI Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra M/Mouse EST Project WashU-HHMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watsn.wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:421034 Seq primer: -28m13 rev2 ET from Amersham High quality sequence stop: 337. Location/Qualifiers 1..346 /organism="Mus musculus" /db_xref="taxon:10090" /clone_image="681330" /clone_lib="Soares mouse NML" /tissue_type="liver" /lab_host="DH10B" /note="vector: pT7n3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'			

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 05:28:30 ; Search time 44.5263 Seconds
(without alignments)
26.934 Million cell updates/sec

Title: US-10-008-355-26

Perfect score: 48
Sequence: 1 TCGNSGSPV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	9	23	AAO15229
2	48	100.0	10	23	AAO15228
3	48	100.0	52	23	AAO15206
4	48	100.0	213	17	AAO15207
5	48	100.0	214	17	AAO15207
6	48	100.0	214	17	AAO15207
7	48	100.0	215	17	AAO15207
8	48	100.0	336	13	AAO15207
9	48	100.0	344	17	AAO15207
10	48	100.0	344	18	AAO15207

11	48	100.0	357	13	AAO15207
12	48	100.0	392	17	AAO15207
13	48	100.0	532	17	AAO15207
14	48	100.0	532	18	AAO15207
15	48	100.0	537	18	AAO15207
16	48	100.0	537	18	AAO15207
17	48	100.0	712	23	AAO15207
18	48	100.0	712	23	AAO15207
19	48	100.0	712	23	AAO15207
20	48	100.0	712	23	AAO15207
21	48	100.0	712	23	AAO15207
22	48	100.0	712	23	AAO15207
23	48	100.0	712	23	AAO15207
24	48	100.0	712	23	AAO15207
25	48	100.0	712	23	AAO15207
26	48	100.0	712	23	AAO15207
27	48	100.0	712	23	AAO15207
28	48	100.0	712	23	AAO15207
29	48	100.0	712	23	AAO15207
30	48	100.0	712	23	AAO15207
31	48	100.0	712	23	AAO15207
32	48	100.0	712	23	AAO15207
33	48	100.0	712	23	AAO15207
34	48	100.0	712	23	AAO15207
35	48	100.0	712	23	AAO15207
36	48	100.0	712	23	AAO15207
37	48	100.0	712	23	AAO15207
38	48	100.0	712	23	AAO15207
39	48	100.0	712	23	AAO15207
40	48	100.0	712	23	AAO15207
41	48	100.0	712	23	AAO15207
42	48	100.0	712	23	AAO15207
43	48	100.0	712	23	AAO15207
44	48	100.0	712	23	AAO15207
45	48	100.0	712	23	AAO15207

ALIGNMENTS

RESULT 1	AAO15229	standard; Peptide: 9 AA.
ID	AAO15229	
XX	AAO15229	
AC	AAO15229	
DT	05-SEP-2002	(first entry)
DE	Porphyrinomas gingivalis dipeptidylpeptidase-7 (DPP-7) active site 2.	
XX	Dipeptidylpeptidase-7; DPP-7; enzyme; amolytic cleavage; active site;	
KW	DPP-7 inhibitor identification; periodontal disease; gingivitis;	
KW	periodontitis.	
XX		
OS	Porphyrinomas gingivalis.	
XX		
PN	WO200238742-A2.	
XX		
PD	16-MAY-2002.	
XX		
PF	08-NOV-2001; 2001WO-US46782.	
XX		
PR	08-NOV-2000; 2000US-246827P.	
XX		
PI	(UYGE-) UNIV GEORGIA RES FOUND INC.	
XX		
PI	Travis J, Potempa JS, Banbula A, Bugno M;	
XX		
DR	WPI; 2002-490075/52.	
XX		
PT	Novel isolated dipeptidylpeptidase useful for identifying inhibitor of	
PT	the dipeptidylpeptidase for protecting an animal from periodontal	
PT	disease caused by Porphyrinomas gingivalis	

XX Claim 4; Page 32; 65pp; English.
PS
XX
CC The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present amino acid sequence represents an active
CC site region of the Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7)
CC enzyme.
XX
SQ Sequence 9 AA;
Query Match 100.0%; Score 48; DB 23; Length 9;
Best Local Similarity 100.0%; Pred. No. 7.8e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGNSGSPV 9
DB 1 TCGNSGSPV 9
RESULT 2
AA015228
ID AA015228 standard; Peptide; 10 AA.
XX
AC AA015228;
XX
DT 05-SEP-2002 (first entry)
XX
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) active site 1.
XX
XX Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage; active site;
XX DPP-7 inhibitor identification; periodontal disease; gingivitis;
XX periodontitis.
XX
XX Porphyromonas gingivalis.
XX OS
XX PN WO200238742-A2.
XX
PD 16-MAY-2002.
XX
PF 08-NOV-2001; 2001WO-US46782.
XX
PR 08-NOV-2000; 2000US-246827P.
XX
PA (UIGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Travis J, Potempa JS, Banbula A, Bugno M;
XX
XX WPI: 2002-490075/52.
DR
XX
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT the dipeptidylpeptidase for protecting an animal from periodontal
PT disease caused by Porphyromonas gingivalis -
XX
PS Claim 5; Page 32; 65pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a

CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present amino acid sequence represents an active
CC site region of the Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7)
CC enzyme.
XX
SQ Sequence 10 AA;
Query Match 100.0%; Score 48; DB 23; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.27;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGNSGSPV 9
DB 1 TCGNSGSPV 9
RESULT 3
AA015206
ID AA015206 standard; Protein; 52 AA.
XX
AC AA015206;
XX
DT 05-SEP-2002 (first entry)
XX
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.
XX
XX Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
XX DPP-7 inhibitor identification; periodontal disease; gingivitis;
XX periodontitis.
XX
XX Porphyromonas gingivalis.
XX OS
XX PN WO200238742-A2.
XX
PD 16-MAY-2002.
XX
PF 08-NOV-2001; 2001WO-US46782.
XX
PR 08-NOV-2000; 2000US-246827P.
XX
PA (UIGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Travis J, Potempa JS, Banbula A, Bugno M;
XX
XX WPI: 2002-490075/52.
DR
XX
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT the dipeptidylpeptidase for protecting an animal from periodontal
PT disease caused by Porphyromonas gingivalis -
XX
PS Example 6; Fig 5; 65pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present amino acid sequence represents the
CC C-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7
CC (DPP-7) enzyme.
XX
SQ Sequence 52 AA;
Query Match 100.0%; Score 48; DB 23; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGNSGSPV 9

DB 1 TCGNSGSPV 9

|||||

RESULT 4

AA015207 standard; Protein: 52 AA.

AA015207;

05-SEP-2002 (first entry)

Staphylococcus aureus V8 endopeptidase C-terminal region.

Dipeptidylpeptidase-7; DPP-7; enzyme: amidolytic cleavage;

DPP-7 inhibitor identification; periodontal disease; gingivitis;

periodontitis; V8 endopeptidase.

Staphylococcus aureus.

MO200238742-A2.

16-MAY-2002.

08-NOV-2001; 2001MO-US46782.

08-NOV-2000; 2000US-246827P.

(UGCE-) UNIV GEORGIA RES FOUND INC.

Travis J, Potempa JS, Banbula A, Bugno M;

WPI; 2002-490075/52.

Novel isolated dipeptidylpeptidase useful for identifying inhibitor of

the dipeptidylpeptidase for protecting an animal from periodontal

disease caused by Porphyromonas gingivalis

Example 6; Fig 5; 65pp; English.

The invention comprises the amino acid and coding sequence of the

Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7

enzyme has amidolytic activity for cleavage of a peptide bond between the

second and third amino acids from the N-terminal end of a target peptide.

The DPP-7 target peptide has an aliphatic or aromatic residue as a

substituent on the alpha-carbon atom of the second amino acid from the

N-terminal end. The DPP-1 protein and DNA sequences of the invention are

useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for

reducing the growth of a bacterium and protecting an animal from a

periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis

or periodontitis). The present amino acid sequence represents the

C-terminal region of the Staphylococcus aureus V8 endopeptidase enzyme.

Sequence 52 AA;

Query Match

Best Local Similarity 100.0%; Score 48; DB 23; Length 52;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9

|||||

1 TCGNSGSPV 9

DB 1 TCGNSGSPV 9

RESULT 5

AA015207 standard; Peptide: 213 AA.

AA015207;

23-MAY-1996 (first entry)

V8 mature protease (aal-213).

XX

Linker peptide: V8 protease; Staphylococcus aureus;

recombinant protein; fusion protein; beta-galactosidase;

Escherichia coli; transposon Tn903;

aminoglycoside 3'-phosphotransferase.

Staphylococcus aureus strain V8 (ATCC 27733).

EP700995-A2.

13-MAR-1996.

06-SEP-1995; 95EP-0306235.

07-NOV-1994; 94JP-0296028.

07-SEP-1994; 94JP-0238595.

(SUNR) SUNTORY LTD.

Ohnaye K, Yabuta M;

WPI; 1996-141021/15.

Prodn. of recombinant polypeptide(s) - using host cells transformed

with a gene coding for the desired polypeptide fused to a protective

polypeptide

Example 8; Fig 14a; 44pp; English.

A fusion protein, V8D (AA015207), has the formula A-L-B-L-C, where A

and C are protective polypeptides, B is Staphylococcus aureus mature

V8 protease lacking the C-terminal repeat region, and L is a linker

host cell, and was cleaved using the E. coli OmpT protease to

yield active V8 protease. Extension of the C-terminal end of the

V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AA015207-44)

1.e. up to residue Phe-215, also resulted in the formation of

inclusion bodies in E. coli. Any further extension gave a soluble

product which exhibited protease activity that repressed growth of

the host cells.

Sequence 213 AA;

Query Match

Best Local Similarity 100.0%; Score 48; DB 17; Length 213;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9

|||||

165 TCGNSGSPV 173

DB 165 TCGNSGSPV 173

RESULT 6

AA015207 standard; Peptide: 214 AA.

AA015207;

23-MAY-1996 (first entry)

V8 mature protease (aal-214).

Linker peptide: V8 protease; Staphylococcus aureus;

recombinant protein; fusion protein; beta-galactosidase;

Escherichia coli; transposon Tn903;

aminoglycoside 3'-phosphotransferase.

Staphylococcus aureus strain V8 (ATCC 27733).

EP700995-A2.

13-MAR-1996.

```
PF 06-SEP-1995; 95EP-0306235.
XX
PR 07-NOV-1994; 94JP-0296028.
PR 07-SEP-1994; 94JP-0238595.
XX
PA (SUNR ) SUNTORY LTD.
XX
PI Ohshuye K, Yabuta M;
XX
DR WPI, 1996-141021/15.
XX
PR Prodn. of recombinant polypeptide(s) - using host cells transformed
PT with a gene coding for the desired polypeptide fused to a protective
PS polypeptide
XX
PS Example 8; Fig 14b; 44pp; English.
XX
CC A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC and C are protective polypeptides, B is Staphylococcus aureus mature
CC V8 protease lacking the C-terminal repeat region, and L is a linker
CC (AAR91032). It was produced as an inclusion body in Escherichia coli
CC host cells, and was cleaved using the E. coli OmpT protease to
CC yield active V8 protease. Extension of the C-terminal end of the
CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)
CC 1.e. up to residue Phe-215, also resulted in the formation of
CC inclusion bodies in E. coli. Any further extension gave a soluble
CC product which exhibited protease activity that repressed growth of
CC the host cells.
XX
SQ Sequence 214 AA:
XX
Query Match 100.0%; Score 48; DB 17; Length 214;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCGNSGSPV 9
DB 165 TCGNSGSPV 173
XX
RESULT 7
AAR91044
ID AAR91044 standard; Peptide: 215 AA.
XX
AC AAR91044;
XX
DT 23-MAY-1996 (first entry)
XX
DE V8 mature protease (aal-215).
XX
XX
XX Linker peptide: V8 protease; Staphylococcus aureus;
XX recombinant protein; fusion protein; beta-galactosidase;
XX Escherichia coli; transposon Tn903;
XX amlnoglucoside 3'-phosphotransferase.
XX
OS Staphylococcus aureus strain V8 (ATCC 27733).
XX
XX EP700995-A2.
XX
XX 13-MAR-1996.
XX
XX 06-SEP-1995; 95EP-0306235.
XX
XX 07-NOV-1994; 94JP-0296028.
XX
XX 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Ohshuye K, Yabuta M;
XX
XX WPI, 1996-141021/15.
XX
XX Prodn. of recombinant polypeptide(s) - using host cells transformed
```

```
PT with a gene coding for the desired polypeptide fused to a protective
PI polypeptide
XX
PS Example 8; Fig 14c; 44pp; English.
XX
CC A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
CC and C are protective polypeptides, B is Staphylococcus aureus mature
CC V8 protease lacking the C-terminal repeat region, and L is a linker
CC (AAR91032). It was produced as an inclusion body in Escherichia coli
CC host cells, and was cleaved using the E. coli OmpT protease to
CC yield active V8 protease. Extension of the C-terminal end of the
CC V8 moiety of the fusion protein by 1, 2 or 3 amino acids (AAR91042-44)
CC 1.e. up to residue Phe-215, also resulted in the formation of
CC inclusion bodies in E. coli. Any further extension gave a soluble
CC product which exhibited protease activity that repressed growth of
CC the host cells.
XX
SQ Sequence 215 AA:
XX
Query Match 100.0%; Score 48; DB 17; Length 215;
Best Local Similarity 100.0%; Pred. No. 6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TCGNSGSPV 9
DB 165 TCGNSGSPV 173
XX
RESULT 8
AAR29644
ID AAR29644 standard; Protein: 336 AA.
XX
AC AAR29644;
XX
DT 11-FEB-1993 (first entry)
XX
DE Protease from S. Aureus.
XX
XX
XX Protease; PCR; amplify; Staphylococcus.
XX
XX Staphylococcus aureus.
XX
XX Key Location/Qualifiers
XX FH 1..68
XX FT /label= Signal_peptide
XX FT 69..336
XX FT /label= Protease
XX
XX JP04211370-A.
XX
XX 03-AUG-1992.
XX
XX 19-FEB-1991; 91JP-0024633.
XX
XX 20-FEB-1990; 90JP-0040398.
XX
XX (SHIO ) SHIONOGI & CO LTD.
XX
XX WPI, 1992-304938/37.
XX
XX N-PSDB; AAQ27988.
XX
XX Novel protease prep. using Bacillus or Saccharomyces host -
XX capable of cleaving peptide bond at carboxyl terminus of glutamic
XX acid residues in polypeptide(s)
XX
XX Disclosure; Page 15-16; 25pp; Japanese.
XX
XX The sequences given in AAR26842 and AAR29644 are proteases which were
XX isolated from Staphylococcus aureus strains. The DNA sequences
XX encoding these proteins were isolated by PCR using the primer
XX sequences given in AAQ27960-86. The protease specifically cleaves
XX the peptide bond at the C-terminus of the glutamic acid residue in
XX polypeptide.
```

```

XX Sequence 336 AA:
SQ
  Query Match 100.0%; Score 48; DB 13; Length 336;
  Best Local Similarity 100.0%; Pred. No. 9.5;
  Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
   |||||
Db 233 TCGNSGSPV 241

RESULT 9
AAR91033
ID AAR91033 standard; Protein; 344 AA.
XX
AC AAR91033;
XX
DT 23-MAY-1996 (first entry)
XX
DE Beta-galactosidase-V8 protease fusion protein.
XX
KM V8 protease; Staphylococcus aureus; recombinant protein;
XX fusion protein; beta-galactosidase; Escherichia coli.
XX
OS Chimeric Escherichia coli;
XX Chimeric Staphylococcus aureus strain V8 (ATCC 27733).
XX
FH Key 1.124
FT Region /note="beta-galactosidase region"
FT 125..344
FT /note="mature V8 protease without the repeat
FT region"

EP700995-A2.
PN
PD 13-MAR-1996.
XX
PF 06-SEP-1995; 95EP-0306235.
XX
PR 07-NOV-1994; 94JP-0296028.
PR 07-SEP-1994; 94JP-0238595.
XX
PA (SUNR ) SUNTORY LTD.
XX
PI Ohsuye K, Yabuta M;
XX WPI; 1996-141021/15.
XX
DR Prodn. of recombinant polypeptide(s) - using host cells transformed
PT with a gene coding for the desired polypeptide fused to a protective
PT polypeptide
XX
PS Example 2; Page 12-13; 44pp; English.
XX
CC Fusion proteins (AAR91033 and AAR91034) were constructed comprising
CC an Escherichia coli beta-galactosidase derivative (protective
CC polypeptide) fused at its C-terminal end to the S. aureus mature
CC V8 protease without or with the repeat region. The constructs
CC were inserted into vector pG97S4DhCR(5)R6, yielding pV8RP7(-) and
CC pV8RP7(+), respectively. Both constructs yielded active protease
CC when expressed in E. coli JM101 transformants.
XX
SQ Sequence 344 AA;

Query Match 100.0%; Score 48; DB 17; Length 344;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
   |||||
Db 289 TCGNSGSPV 297

```

```

RESULT 10
AAM22218
ID AAM22218 standard; Protein; 344 AA.
XX
AC AAM22218;
XX
DT 11-SEP-1997 (first entry)
XX
DE Protein encoded by pV8RP7(-) construct.
XX
KM Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;
XX truncation; wild type; PCR; polymerase chain reaction; amplification;
XX proteolytic activity; fusion protein; beta-galactosidase; urea.
XX
OS Chimeric - Escherichia coli.
XX Chimeric - Staphylococcus aureus.
XX
FH Key 1.100
FT Region /note="E. coli beta-galactosidase portion"
FT 101..120
FT /note="synthetic R6 linker"
FT 125..344
FT /note="truncated S. aureus V8 protease portion"

EP745669-A2.
PN
PD 04-DEC-1996.
XX
PF 31-MAY-1996; 96EP-0303939.
XX
PR 02-JUN-1995; 95JP-0170086.
XX
PA (SUNR ) SUNTORY LTD.
XX
PI Ohsuye K, Yabuta M;
XX WPI; 1997-013693/02.
XX
DR Staphylococcus aureus V8 protease mutants - with increased
PT resistance to denaturation
PT
XX
PS Claim 2; Page 13-14; 42pp; English.
XX
CC The invention relates to new mutant Staphylococcus aureus V8 proteases
CC which have enzyme activity even under environmental conditions which
CC promote protein denaturation. The mutants are based on 3 truncated V8
CC proteases lacking 48 (AAM22218), 56 (AAM22219) or 53 (AAM22220) amino
CC acids from the C-terminal of the wild type protease. The mutants also
CC contain amino acid substitutions, especially D44E, N71S and/or R147K.
CC The protein sequence shown here represents a chimeric protein
CC comprising a truncated Staphylococcus aureus V8 protease lacking the
CC prepro and C-terminal 48 amino acids linked, via a synthetic linker,
CC downstream of the E. coli beta-galactosidase. The S. aureus portion of
CC the chimera was amplified by the primers AAT73254-5 from wild type
CC sequence. The coding sequence was then used to generate mutants of a
CC V8 protease which retain their levels of activity in the presence of a
CC higher concentration of protein denaturant e.g. 5 M urea.
XX
SQ Sequence 344 AA;

Query Match 100.0%; Score 48; DB 18; Length 344;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
   |||||
Db 289 TCGNSGSPV 297

```

RESULT 11

AA	R26842	standard; Protein; 357 AA.
ID	AA	R26842
AC	AA	R26842;
XX	AA	R26842;
DT	11-FEB-1993	(first entry)
XX	DE	Protease from S. Aureus ATCC12600.
XX	XX	Protease; PCR; amplify; Staphylococcus.
OS	Staphylococcus aureus.	
XX	XX	
EH	Key	Location/Qualifiers
FT	Peptide	1..68
FT	Protein	/label= Signal_peptide
FT		69..358
FT		/label= Protease
XX	XX	
PN	JP04211370-A.	
XX	XX	
PD	03-AUG-1992.	
XX	XX	
PE	19-FEB-1991;	91JP-0024633.
XX	XX	
PR	20-FEB-1990;	90JP-0040398.
XX	XX	
PA	(SHIO) SHIONOGI & CO LTD.	
XX	XX	
DR	WPI: 1992-304938/37.	
DR	N-PSDB; AAQ27987.	
XX	XX	
PT	Novel protease prep'd. using Bacillus or Saccharomyces host	
PT	capable of cleaving peptide bond at carboxyl terminus of glutamic	
PT	acid residues in polypeptide(s)	
XX	XX	
PS	Disclosure: Page 13-15; 25pp; Japanese.	
XX	XX	
CC	The sequences given in AAR26842 and AAR29644 are proteases which were	
CC	isolated from Staphylococcus aureus strains. The DNA sequences	
CC	encoding these proteins were isolated by PCR using the primer	
CC	sequences given in AAQ27960-86. The protease specifically cleaves	
CC	the peptide bond at the C-terminus of the glutamic acid residue in	
CC	polypeptide.	
XX	XX	
SQ	Sequence	357 AA;
XX	XX	
Query Match	100.0%;	Score 48; DB 13; Length 357;
Best Local Similarity	100.0%;	Pred. No. 10;
Matches	9; Conservative	0; Mismatches 0; Indels 0; Gaps 0.
OY	1 TCGNGSPV 9	
DB	233 TCGNGSPV 241	
XX	XX	
RESULT 12	AAAR91034	
ID	AAAR91034	standard; Protein; 392 AA.
AC	AAAR91034;	
XX	XX	
DT	23-MAY-1996	(first entry)
XX	XX	
DE	Beta-galactosidase-V8 protease fusion protein.	
XX	XX	
V8	protease; Staphylococcus aureus; recombinant protein;	
RW	fusion protein; beta-galactosidase; Escherichia coli.	
XX	XX	
OS	Chimeric Escherichia coli;	
OS	Chimeric Staphylococcus aureus strain V8 (ATCC 27733).	
XX	XX	
Key	Location/Qualifiers	

	FT	Region	1..124	/note= "beta-galactosidase region"	
	FT		123..392		
	FT	Region	/note= "mature V8 protease including the repeat region"		
	XX	EP700995-A2.			
	PN	13-MAR-1996.			
	PD				
	PE	06-SEP-1995;	95EP-0306235.		
	XX				
	PR	07-MOV-1994;	94JP-0296028.		
	PK	07-SEP-1994;	94JP-0238595.		
	XX	(SUNR) SUNTORY LTD.			
	PA	Ohsuye K, Yabuta M;			
	PI				
	DR	WPT; 1996-141021/15.			
	XX				
	PP	Prodn. of recombinant polypeptide(s) - using host cells transformed with a gene coding for the desired polypeptide fused to a protective polypeptide			
	PS	Example 2; Page 14-15; 44pp; English.			
	CC	Fusion proteins (AAR91033 and AAR91034) were constructed comprising an Escherichia coli beta-galactosidase derivative (protective polypeptide) fused at its C-terminal end to the S. aureus mature V8 protease without or with the repeat region. The constructs were inserted into vector pG9/S4DHCT(G)H6, yielding PV8RP(+) and PV8PRT(+), respectively. Both constructs yielded active protease when expressed in E. coli JM101 transformants.			
	SQ	Sequence 392 AA:			
	Query Match	100.0%; Score 48; DB 17; Length 392; Best Local Similarity 100.0%; Pred. No. 11; Matches 9; Conservative % 0; Mismatches 0; Indels 0; Gaps 0;			
	CY	1 TGGNSGPSV 9 			
	DB	289 TGGNSGPSV 297			
	RESULT_13				
	AAR91035				
	ID	AAR91035 standard; Protein: 532 AA.			
	XX				
	AC	AAR91035;			
	DT	23-MAY-1996 (first entry)			
	DE	Recombinant V8 protease V8D fusion protein.			
	XX				
	RN	Linker peptide; V8 protease; Staphylococcus aureus; recombinant protein; fusion protein; beta-galactosidase; Escherichia coli; transposon Th903; aminoglycoside 3'-phosphotransferase.			
	KX				
	OS	Chimeric Escherichia coli;			
	OS	Chimeric synthetic;			
	OS	Chimeric staphylococcus aureus strain V8 (ATCC 27733); Chimeric transposon Th903.			
	XX				
	EH	Key Location/Qualifiers			
	FT	Region 1..100			
	FT	Region /note= "beta-galactosidase region"			
	FT	Region 101..120			
	FT	Region /note= "R6 linker"			
	FT	Cleavage-site 104..105			
	TT	/note= "cleavage site for OmpT protease"			

```
FT Region 125..335
FT /note= "V8 protease region"
FT Region 336..356
FT /note= "R6 linker"
FT Cleavage-site 339..340
FT /note= "cleavage site for Ompt protease"
FT Region 307..532
FT /note= "aminoglycoside 3'phosphotransferase
FT region"
XX
XX EP700995-A2.
XX
XX 13-MAR-1996.
XX
XX 06-SEP-1995; 95EP-0306235.
XX
XX 07-NOV-1994; 94JP-0296028.
XX 07-SEP-1994; 94JP-0238595.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Ohshuye K, Yabuta M;
XX
XX WPI: 1996-141021/15.
XX
XX Prodn. of recombinant polypeptide(s) - using host cells transformed
XX with a gene coding for the desired polypeptide fused to a protective
XX polypeptide
XX
XX Example 3; Page 16-18; 44pp; English.
XX
XX A fusion protein, V8D (AAR91035), has the formula A-L-B-L-C, where A
XX and C are protective polypeptides (derived from Escherichia coli
XX beta-galactosidase and and Tn903 aminoglycoside 3'-phosphotransferase,
XX respectively), B is Staphylococcus aureus mature V8 protease
XX lacking the C-terminal repeat region, and L is a linker peptide
XX (AAR91032). The fusion protein is expressed in inactive form in E.
XX coli. It is then recovered, solubilised and cleaved at the linker
XX peptide regions with a protease intrinsic to the host cells, i.e.
XX Ompt protease, to allow recovery of V8 protease.
XX
XX Sequence 532 AA:
SQ
Query Match 100.0%; Score 48; DB 17; Length 532;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGGNSGSPV 9
DB 289 TGGNSGSPV 297
RESULT 14
AAW22219
ID AAW22219 standard; Protein: 532 AA.
XX
XX AAW22219;
XX
XX 11-SEP-1997 (first entry)
XX
XX Protein encoded by pV8D construct.
XX
XX Mutant; Staphylococcus aureus; V8 protease; enzyme: denaturation; primer;
XX truncation: wild type; PCR; polymerase chain reaction; amplification;
XX proteolytic activity; fusion protein; beta-galactosidase; urea.
XX
XX Chimeric - Escherichia coli.
XX Chimeric - Staphylococcus aureus.
XX
XX Key Location/Qualifiers
XX Region 1..100
XX /note= "E. coli beta-galactosidase portion"
XX Region 101..124
```

```
FT Region /note= "R6 linker sequence"
FT Region 125..336
FT /note= "truncated S. aureus V8 protease portion"
FT Region 337..360
FT /note= "R6 linker sequence"
FT Region 361..532
FT /note= "aminoglycoside 3'-phosphotransferase portion"
XX
XX EP745669-A2.
XX
XX 04-DEC-1996.
XX
XX 31-MAY-1996; 96EP-0303939.
XX
XX 02-JUN-1995; 95JP-0170086.
XX
XX (SUNR ) SUNTORY LTD.
XX
XX Ohshuye K, Yabuta M;
XX
XX WPI: 1997-013693/02.
XX
XX Staphylococcus aureus V8 protease mutants - with increased
XX resistance to denaturation
XX
XX Claim 7; Page 16-17; 42pp; English.
XX
XX The invention relates to new mutant Staphylococcus aureus V8 proteases
XX which have enzyme activity even under environmental conditions which
XX promote protein denaturation. The mutants are based on 3 truncated V8
XX proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
XX acids from the C-terminal of the wild type protease. The mutants also
XX contain amino acid substitutions, especially D44E, N71S and/or R147K.
XX The protein sequence shown here represents a chimeric protein
XX comprising a truncated Staphylococcus aureus V8 protease lacking the
XX prepro and C-terminal 56 amino acids linked, via a synthetic R6 linker,
XX downstream of the E. coli beta-galactosidase. Also included downstream of
XX the V8 protease fragment is a second R6 linker and a fragment of the
XX aminoglycoside 3'-phosphotransferase protein. The chimeric sequence
XX was generated by restriction digestion and ligation from the V8KPT(-)
XX sequence (see AAW22218) by using a natural EcoRV site which removed a
XX further 8 amino acid from the C-terminus. This truncated V8 protease,
XX designated V8D, retains its level of activity in the presence of a
XX higher concentration of protein denaturant e.g. 5 M urea.
XX
XX Sequence 532 AA:
SQ
Query Match 100.0%; Score 48; DB 18; Length 532;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TGGNSGSPV 9
DB 289 TGGNSGSPV 297
RESULT 15
AAW22220
ID AAW22220 standard; Protein: 537 AA.
XX
XX AAW22220;
XX
XX 11-SEP-1997 (first entry)
XX
XX Protein encoded by pV8F construct.
XX
XX Mutant; Staphylococcus aureus; V8 protease; enzyme: denaturation; primer;
XX truncation: wild type; PCR; polymerase chain reaction; amplification;
XX proteolytic activity; fusion protein; beta-galactosidase; urea.
XX
XX Chimeric - Escherichia coli.
XX Chimeric - Staphylococcus aureus.
```

```

FH Key Location/Qualifiers
FT Region 1..100
FT /note="E. coli beta-galactosidase"
FT Region 101..124
FT /note="R6 linker sequence"
FT Region 125..339
FT /note="truncated S. aureus V8 protease"
FT Region 342..365
FT /note="R6 linker sequence"
FT Region 366..537
FT /note="aminoglucoside 3'-phosphotransferase"
PN EP/45669-A2.
XX 04-DEC-1996.
XX
XX 31-MAY-1996; 96EP-0303939.
XX
XX 02-JUN-1995; 95JP-0170086.
XX
XX (SUNR ) SUNTORX LTD.
XX
XX PI Ohshye K, Yabuta M;
XX WPI; 1997-013693/02.
XX
XX Staphylococcus aureus V8 protease mutants - with increased
XX resistance to denaturation
XX
XX Claim 8, Page 19-20; 42pp; English.
XX
XX The invention relates to new mutant Staphylococcus aureus V8 proteases
XX which have enzyme activity even under environmental conditions which
XX promote protein denaturation. The mutants are based on 3 truncated V8
XX proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
XX acids from the C-terminal of the wild type protease. The mutants also
XX contain amino acid substitutions, especially D44E, N71S and/or R147K.
XX The protein sequence shown here represents a chimeric protein
XX comprising a truncated Staphylococcus aureus V8 protease lacking the
XX prepro and C-terminal 53 amino acids linked, via a synthetic R6 linker,
XX downstream of the E. coli beta-galactosidase. Also included downstream of
XX the V8 protease fragment is a second R6 linker and a fragment of the
XX aminoglucoside 3'-phosphotransferase protein. The chimeric sequence
XX was generated by restriction digestion and ligation from the V8RPF(-)
XX sequence (see AAW22218) by using a natural EcoRV site which removed a
XX further 8 amino acid from the C-terminus. This truncated V8 protease,
XX designated V8F, retains its level of activity in the presence of a
XX higher concentration of protein denaturant e.g. 5 M urea.
XX
XX SQ Sequence 537 AA:

```

```

Query Match 100.0%; Score 48; DB 18; Length 537;
Best Local Similarity 100.0%; Pred. NO. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCGNSGSPV 9
| | | | | | | |
DB 289 TCGNSGSPV 297

```

Search completed: December 20, 2002, 11:24:53
Job time : 45.5263 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 11:25:03 ; Search time 11.3684 Seconds
(without alignments)
13.271 Million cell updates/sec

Title: US-10-008-355-26

Perfect score: 48

Sequence: 1 TGNSSGSPV 9

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PC1_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PC1US_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	9	US-10-008-355-26	Sequence 26, Appl
2	48	100.0	10	US-10-008-355-25	Sequence 25, Appl
3	48	100.0	52	US-10-008-355-3	Sequence 3, Appl
4	48	100.0	52	US-10-008-355-4	Sequence 4, Appl
5	48	100.0	699	US-10-008-355-8	Sequence 8, Appl
6	48	100.0	712	US-10-008-355-2	Sequence 2, Appl
7	48	100.0	716	US-10-008-355-7	Sequence 7, Appl
8	48	100.0	720	US-10-008-355-9	Sequence 9, Appl
9	48	100.0	732	US-10-008-355-6	Sequence 6, Appl
10	44	91.7	734	US-10-008-355-5	Sequence 5, Appl
11	37	77.1	74	US-09-864-761-48349	Sequence 48349, A
12	37	77.1	240	US-09-815-242-5351	Sequence 5351, Ap
13	37	77.1	254	US-09-815-242-12277	Sequence 12277, A
14	33	68.8	40	US-09-864-761-34778	Sequence 34778, A
15	33	68.8	74	US-09-864-761-42729	Sequence 42729, A
16	33	68.8	223	US-10-003-152-4	Sequence 4, Appl
17	33	68.8	1070	US-09-735-367B-6	Sequence 6, Appl
18	33	68.8	1981	US-09-928-457-38	Sequence 38, Appl
19	33	68.8	2005	US-09-735-367B-3	Sequence 3, Appl

20	33	68.8	2063	10	US-09-735-367B-2	Sequence 2, Appl
21	32	66.7	43	10	US-09-864-761-40613	Sequence 40613, A
22	32	66.7	94	10	US-09-864-761-34548	Sequence 34548, A
23	32	66.7	97	10	US-09-734-017A-40	Sequence 40, Appl
24	32	66.7	180	9	US-10-023-182-8	Sequence 8, Appl
25	32	66.7	180	9	US-09-751-798-8	Sequence 8, Appl
26	32	66.7	230	10	US-09-815-242-13533	Sequence 13533, A
27	32	66.7	351	10	US-09-853-625B-16	Sequence 16, Appl
28	32	66.7	371	10	US-09-739-861A-1	Sequence 1, Appl
29	32	66.7	371	10	US-09-795-583-1	Sequence 1, Appl
30	32	66.7	386	10	US-09-739-861A-5	Sequence 5, Appl
31	32	66.7	386	10	US-09-795-583-5	Sequence 5, Appl
32	32	66.7	397	10	US-09-821-883-27	Sequence 27, Appl
33	32	66.7	422	10	US-09-815-242-11329	Sequence 11329, A
34	32	66.7	672	10	US-09-858-754-2	Sequence 2, Appl
35	32	66.7	701	10	US-09-815-242-13411	Sequence 13411, A
36	32	66.7	758	10	US-09-801-368-224	Sequence 224, App
37	32	66.7	1034	10	US-09-858-754-5	Sequence 5, Appl
38	32	66.7	1302	12	US-10-000-864-2	Sequence 2, Appl
39	32	66.7	1493	10	US-09-858-754-3	Sequence 3, Appl
40	32	66.7	1493	12	US-10-000-864-8	Sequence 8, Appl
41	32	66.7	1770	10	US-09-841-132-444	Sequence 444, App
42	32	66.7	2502	10	US-09-772-316-1	Sequence 1, Appl
43	32	66.7	2905	8	US-08-424-550B-401	Sequence 401, App
44	31	64.6	57	10	US-09-864-761-43752	Sequence 43752, A
45	31	64.6	136	10	US-09-864-761-48257	Sequence 48257, A

ALIGNMENTS

RESULT 1
US-10-008-355-26
Sequence 26, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylleptidases And Methods Of Use
FILE REFERENCE: 235,00440101
CURRENT APPLICATION NUMBER: US/10/008,355
PRIORITY FILING DATE: 2001-11-08
PRIORITY APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.0
SEQ ID NO 26
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus sequence
US-10-008-355-26

Query Match 100.0%; Score 48; DB 9; Length 9;
Best Local Similarity 100.0%; Pred. No. 8.8e+04;
Matches: 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGNSSGSPV 9
|||||||
DB 1 TGNSSGSPV 9

RESULT 2
US-10-008-355-25
Sequence 25, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylleptidases And Methods Of Use

FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.0
SEQ ID NO 25
LENGTH: 10
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus sequence
US-10-008-355-25

Query Match 100.0%; Score 48; DB 9; Length 10;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||

DB 1 TCGNSGSPV 9

AD

RESULT 3
US-10-008-355-3
Sequence 3, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 52
TYPE: PRT
ORGANISM: Porphyromonas gingivalis
US-10-008-355-3

Query Match 100.0%; Score 48; DB 9; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||

DB 1 TCGNSGSPV 9

AD

RESULT 4
US-10-008-355-4
Sequence 4, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 52

TYPE: PRT
ORGANISM: Staphylococcus aureus
US-10-008-355-4

Query Match 100.0%; Score 48; DB 9; Length 52;
Best Local Similarity 100.0%; Pred. No. 0.064;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||

DB 1 TCGNSGSPV 9

AD

RESULT 5
US-10-008-355-8
Sequence 8, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.0
SEQ ID NO 8
LENGTH: 699
TYPE: PRT
ORGANISM: Porphyromonas gingivalis
US-10-008-355-8

Query Match 100.0%; Score 48; DB 9; Length 699;
Best Local Similarity 100.0%; Pred. No. 0.85;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||

DB 644 TCGNSGSPV 652

AD

RESULT 6
US-10-008-355-2
Sequence 2, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potempa, Jan S
APPLICANT: Banbula, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235.00440101
CURRENT APPLICATION NUMBER: US/10/008,355
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: US 60/246,827
PRIOR FILING DATE: 2000-11-08
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 712
TYPE: PRT
ORGANISM: Porphyromonas gingivalis
US-10-008-355-2

Query Match 100.0%; Score 48; DB 9; Length 712;
Best Local Similarity 100.0%; Pred. No. 0.86;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
|||||

DB 644 TCGNSGSPV 652

RESULT 7
US-10-008-355-7
; Sequence 7, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-008-355-7

AD

Query Match 100.0%; Score 48; DB 9; Length 716;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNCGSPV 9
|||||
DB 647 TCGNCGSPV 655

RESULT 8
US-10-008-355-9
; Sequence 9, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-9

AD

Query Match 100.0%; Score 48; DB 9; Length 720;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNCGSPV 9
|||||
DB 651 TCGNCGSPV 659

AD

RESULT 9
US-10-008-355-6
; Sequence 6, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use

; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-6

Query Match 100.0%; Score 48; DB 9; Length 732;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNCGSPV 9
|||||
DB 662 TCGNCGSPV 670

A

RESULT 10
US-10-008-355-5
; Sequence 5, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Shewanella putrefaciens
US-10-008-355-5

Query Match 91.7%; Score 44; DB 9; Length 734;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGNCGSP 8
|||||
DB 665 TCGNCGSP 672

RESULT 11
US-09-864-48349
; Sequence 48349, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03

```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48349
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL121716.16
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.9
; OTHER INFORMATION: EST_HUMAN HIT: BE002805.1, EVALUE 8.00e-39
; OTHER INFORMATION: SWISSPROT HIT: Q9ZKD2, EVALUE 3.60e+00
US-09-864-761-48349

Query Match          77.1%; Score 37; DB 10; Length 74;
Best Local Similarity 75.0%; Pred. No. 6.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 GNGSGSPV 9
        11:|||||
DB      52 GCHSGSPI 59

RESULT 12
US-09-815-242-5351
; Sequence 5351, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

```

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12277
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12277

Query Match          77.1%; Score 37; DB 10; Length 254;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 GNGSGSPV 9
        |||||
DB      191 GNGSGSPV 197

RESULT 13
US-09-815-242-12277
; Sequence 12277, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12277
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12277
```

Db 205 GNGSPV 211

RESULT 14

```
US-09-864-761-34778
; Sequence 34778, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmics-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34778
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010092.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.94
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1
; OTHER INFORMATION: EST_HUMAN HIT: BE939264.1, EVALUATE 7.00e-04
; US-09-864-761-34778
```

Query Match 68.8%; Score 33; DB 10; Length 40;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GNGSPV 9
Db 11 GNGSPV 18

RESULT 15

```
US-09-864-761-42729
; Sequence 42729, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmics-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 42729
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF176315.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.93
; OTHER INFORMATION: EST_HUMAN HIT: T92522.1, EVALUATE 5.00e-24
; OTHER INFORMATION: SWISSPROT HIT: P23471, EVALUATE 1.20e+00
; US-09-864-761-42729
```

Query Match 68.8%; Score 33; DB 10; Length 74;
 Best Local Similarity 66.7%; Pred. No. 28;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGNSGSPV 9
 | | | | |
 Db 6 TAGTSGNPV 14

Search completed: December 20, 2002, 12:17:27
 Job time : 12.3684 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 08:33:52 ; Search time 24.1579 Seconds
(Without alignments)
35.815 Million cell updates/sec

Title: US-10-008-355-26

Perfect score: 48

Sequence: 1 TCGNSGSPV 9

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	48	100.0	336	1	PR5ASK	glutamyl endopepti
2	48	100.0	342	2	G89873	hypothetical prote
3	48	100.0	357	2	S21758	glutamic acid-spec
4	48	100.0	716	2	G82627	hypothetical prote
5	42	87.5	169	2	A12436	hypothetical prote
6	41	85.4	284	2	S21440	serine proteinase
7	40	83.3	315	2	E71729	proteinase DO (htr
8	40	83.3	361	2	B56940	integral membrane
9	40	83.3	363	2	A56940	integral membrane
10	40	83.3	497	2	B97729	heat shock protein
11	39	81.2	996	2	T50186	hypothetical signa
12	39	81.2	997	2	T39521	hypothetical prote
13	39	81.2	997	2	S63064	probable membrane
14	38	79.2	476	2	B82414	leucine aminopepti
15	37	77.1	182	2	T49812	hypothetical prote
16	37	77.1	235	2	D89967	serine proteinase
17	37	77.1	239	2	B89967	serine proteinase
18	37	77.1	240	2	C89967	serine proteinase
19	37	77.1	316	2	A45134	endopeptidase (EC
20	37	77.1	401	2	A48423	engorged homeodom
21	37	77.1	449	2	B71265	probable Mg2+ tran
22	37	77.1	590	2	A81411	autolysin, N-acety
23	37	77.1	817	2	T21336	hypothetical prote
24	37	77.1	1254	2	T47141	hypothetical prote
25	37	77.1	1306	2	A70934	hypothetical glyci
26	37	77.1	1777	2	T34369	hypothetical prote
27	36	75.0	320	2	UC1311	cell protein precu
28	36	75.0	374	2	T09111	probable magnesium
29	36	75.0	379	2	S23573	pectin lyase (EC 4

30	36	75.0	411	2	T15209	hypothetical prote
31	36	75.0	451	2	A82172	magnesium transpor
32	36	75.0	591	2	T25636	hypothetical prote
33	36	75.0	1129	2	T25635	hypothetical prote
34	36	75.0	1175	2	T25634	hypothetical prote
35	36	75.0	1283	2	T49804	hypothetical prote
36	36	75.0	1711	2	T21432	hypothetical prote
37	35	72.9	108	2	T10377	hypothetical prote
38	35	72.9	463	2	A61542	wall associated pr
39	35	72.9	521	2	S62794	probable lipoprote
40	35	72.9	531	2	S62796	probable lipoprote
41	35	72.9	584	2	S55106	probable membrane
42	35	72.9	656	2	C83479	probable proteinase
43	35	72.9	783	2	E70824	hypothetical glyci
44	35	72.9	1655	2	T13998	gene mastermind pr
45	35	72.9	1714	1	S18644	multifunctional am

ALIGNMENTS

```
RESULT 1
PR5ASK
glutamyl endopeptidase (EC 3.4.21.19) precursor - Staphylococcus aureus
N:Alternate names: staphylococcal serine proteinase
C:Species: Staphylococcus aureus
C:Date: 04-Dec-1986 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1999
C:Accession: A26812; A00966
R:Carmona, C.; Gray, G.L.
Nucleic Acids Res. 15, 6757, 1987
A:Title: Nucleotide sequence of the serine proteinase gene of Staphylococcus aureus, st
A:Reference number: A26812; MUID:87316953; PMID:3306605
A:Accession: A26812
A:Molecule type: DNA
A:Residues: 1-336 <CAR>
A:Cross-references: EMBL:Y00356; NID:946686; PIDN:CA68434.1; PID:946687
A:Reference number: A23824; MUID:78212487; PMID:96922
A:Accession: A00966
A:Molecule type: Protein
A:Residues: 69-108;110-124;126-144,'D',146-192,'T',194-228,'N',230-258,'Q',260,'D',26
Can. J. Biochem. 56, 534-544, 1978
A:Title: The primary structure of staphylococcal protease.
A:Reference number: A23824; MUID:78212487; PMID:96922
A:Accession: A00966
A:Molecule type: Protein
A:Residues: 69-108;110-124;126-144,'D',146-192,'T',194-228,'N',230-258,'Q',260,'D',26
R:Draper, G.R.
C:Experimental source: strain V8
C:Comment: This extracellular proteolytic enzyme preferentially cleaves the peptide b
C:Comment: This enzyme may be distantly related to the trypsin-type serine proteinase
C:Superfamily: staphylococcal serine proteinase
C:Keywords: hydrolase; serine proteinase
F:69-336/Product: staphylococcal serine proteinase #status experimental <ACT>
F:119,161/Active site: His, Asp #status predicted
F:237/Active site: Ser #status experimental

Query Match 100.0%; Score 48; DB 1; Length 336;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TCGNSGSPV 9
Db 233 TCGNSGSPV 241

RESULT 2
G89873
hypothetical protein ssps [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: G89873
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Gai, L.; O
ma, A.; Mizutani-Oi, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
```

A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: G89873
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-342 <KUR>
A:Cross-references: GB:BA000018; PID:g13700850; PIDN:BAB42146.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: ssbA
C:Superfamily: staphylococcal serine proteinase

Query Match 100.0%; Score 48; DB 2; Length 342;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCGNSGSPV 9
Db 233 TCGNSGSPV 241

RESULT 3
S21758
glutamic acid-specific endopeptidase - *Staphylococcus aureus*
C:Species: *Staphylococcus aureus*
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C:Accession: S21758
R:Yoshikawa, K.; Tsunuki, H.; Fujiwara, T.; Nakamura, E.; Iwamoto, H.; Matsumoto, K.; Shi-
Biotchim. Biophys. Acta 1121, 221-228, 1992
A:Title: Purification, characterization and gene cloning of a novel glutamic acid-specific
A:Reference number: S21758; MUID:92287954; PMID:1599945
A:Accession: S21758
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <YOS>
A:Cross-references: GB:D00730; NID:g216970; PIDN:BA00630.1; PID:g216971
C:Superfamily: staphylococcal serine proteinase

Query Match 100.0%; Score 48; DB 2; Length 357;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCGNSGSPV 9
Db 233 TCGNSGSPV 241

RESULT 4
G82627
hypothetical protein XF1887 [imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82627
R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen-
tation 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82627
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-716 <STI>
A:Cross-references: GB:AE004008; GB:AE003849; NID:g9106961; PIDN:AF84693.1; GSPDB:GN001
R:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A-
Briotes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carret, H-
as-Melo, E.; Docena, C.; El-Dorriy, H.; Facinican, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm-
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigt-
chado, M.A.; Medeira, A.M.B.N.; Medeira, H.M.F.; Marinho, C.L.; Marques, M.V.; Martins, B-
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv-
M.; Tsubako, M.H.; Valleda, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF1887

Query Match 100.0%; Score 48; DB 2; Length 716;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCGNSGSPV 9
Db 647 TCGNSGSPV 655

RESULT 5
A12436
hypothetical protein alr5049 [imported] - *Nostoc* sp. (strain PCC 7120)
C:Species: *Nostoc* sp.
A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: A12436
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kurlitz, T.; Sasamoto, S.; Watanabe, A.; Irigu-
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759640
A:Accession: A12436
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-169 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA876748.1; PID:g17134187; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr5049

Query Match 87.5%; Score 42; DB 2; Length 169;
Best Local Similarity 88.9%; Pred. No. 5.3;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCGNSGSPV 9
Db 109 TCGNSASPV 117

RESULT 6
S25140
serine proteinase homolog - *Enterococcus faecalis*
C:Species: *Enterococcus faecalis*
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-2000
C:Accession: S25140
R:Su, Y.A.; Clewell, D.B.
submitted to the EMBL Data Library, June 1992
A:Description: A gene (SPR) downstream of *gelE* of *Enterococcus faecalis* OG1-10 resem-
A:Reference number: S25140
A:Accession: S25140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-284 <SOY>
A:Cross-references: EMBL:212296; NID:g43337; PIDN:CA78168.1; PID:g43338
C:Superfamily: staphylococcal serine proteinase

Query Match 85.4%; Score 41; DB 2; Length 284;
Best Local Similarity 77.8%; Pred. No. 13;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCGNSGSPV 9
Db 224 TCGGSGSPI 232

RESULT 7
E71729
proteinase DO (htira) RP186 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C:Accession: E71729
R:Andersson, S.G.E.; ZomorodiPour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alsmark, U.
Nature 386, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: E71729
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-315 <AND>
A:Cross-references: GB:AJ235270; GB:AJ235269; NID:g3860572; PIDN:CAA14652.1; PID:e134249
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: htira; RP186

Query Match 83.3%; Score 40; DB 2; Length 315;
Best Local Similarity 88.9%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
||| |||||
Db 188 TCGAGSGSPV 196

RESULT 8
B56940
Integral membrane protein TGN38A - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995
C:Accession: B56940
R:Kasai, K.; Takahashi, S.; Murakami, K.; Nakayama, K.
J. Biol. Chem. 270, 14471-14476, 1995
A:Title: Strain-specific presence of two TGN38 isoforms and absence of TGN41 in mouse.
A:Reference number: A56940; MUID:95301533; PMID:7540170
A:Accession: B56940
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-361 <KAS>
A:Cross-references: GB:ID50032

Query Match 83.3%; Score 40; DB 2; Length 361;
Best Local Similarity 87.5%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSP 8
|||||
Db 138 TCGNSGSP 145

RESULT 9
A56940
Integral membrane protein TGN38B - mouse
C:Species: Mus musculus (house mouse)
C:Date: 19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995
C:Accession: A56940
R:Kasai, K.; Takahashi, S.; Murakami, K.; Nakayama, K.
J. Biol. Chem. 270, 14471-14476, 1995
A:Title: Strain-specific presence of two TGN38 isoforms and absence of TGN41 in mouse.
A:Reference number: A56940; MUID:95301533; PMID:7540170
A:Accession: A56940
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-363 <KAS>
A:Cross-references: GB:ID50031

Query Match 83.3%; Score 40; DB 2; Length 363;
Best Local Similarity 87.5%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSP 8
|||||
Db 140 TCGNSGSP 147

RESULT 10
B97729
heat shock proteinase (EC 3.4.21.-) [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: B97729
R:ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: B97729
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-497 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL02772.1; PID:g15619287; GSPDB:GN00173
C:Genetics:
A:Gene: RC0234
C:Keywords: hydrolase; serine proteinase

Query Match 83.3%; Score 40; DB 2; Length 497;
Best Local Similarity 88.9%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
||| |||||
Db 188 TCGAGSGSPV 196

RESULT 11
T50186
Hypothetical signaling-associated PDZ domain containing protein [imported] - fission
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T50186
R:Badcock, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 2000
A:Reference number: Z25044
A:Accession: T50186
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-996 <BAD>
A:Cross-references: EMBL:AL138854; PIDN:CA872237.1; GSPDB:GN00066; SPDB:SPAC2363.12c
A:Experimental source: strain 972h(-); cosmid c2363
C:Genetics:
A:Gene: SPDB:SPAC2363.12c
A:Map position: 1

Query Match 81.2%; Score 39; DB 2; Length 996;
Best Local Similarity 77.8%; Pred. No. 95;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
:|||||
Db 212 SGGSGSPV 220

RESULT 12
T39521
Hypothetical protein SPBC1685.05 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T39521
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Hildbert, H.; Duesterhoeft, A.
submitted to the EMBL Data Library, July 1998
A:Reference number: Z21861
A:Accession: T39521
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-997 <WOO>
A:Cross-references: EMBL:AL031154; PIDN:CAA20053.1; GSPDB:GN00067; SPDB:SPBC1685.05
A:Experimental source: strain 972h.; cosmid c1685
C:Genetics:
A:Gene: SPDB:SPBC1685.05
A:Map position: 2

Query Match 81.2%; Score 39; DB 2; Length 997;
Best Local Similarity 77.8%; Pred. No. 96;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
DB 227 SGGSSGSPV 235

RESULT 13
S63064
probable membrane protein YNL123w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein NI897
C:Species: Saccharomyces cerevisiae

C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 19-Apr-2002
C:Accession: S63064; S67335

R:De Antonl, A.; D Angelo, M.; Dal Pero, F.; Sartorello, F.; Pandolfo, D.; Lanfranchi, G.
submitted to the Protein Sequence Database, April 1996

A:Reference number: S63047
A:Accession: S63064

A:Molecule type: DNA

A:Residues: 1-997 <DEA>

A:Cross-references: EMBL:Z71399; NID:g1302053; PID:e239802; PID:g1302054; MIPS:YNL123w

A:Experimental source: strain S288C

R:d'Antoni, A.; d'Angelo, M.; dal Pero, F.; Sartorello, F.; Pandolfo, D.; Lanfranchi, G.
submitted to the EMBL Data Library, February 1996

A:Description: The DNA sequence of cosmid 14-13b from chromosome XIV of Saccharomyces cerevisiae
A:Reference number: S67327

A:Accession: S67335

A:Molecule type: DNA

A:Residues: 1-997 <DAN>

A:Cross-references: EMBL:Z69382; NID:g1183941; PID:e221825; PID:g1183950

C:Genetics:
A:Cross-references: SGD:S0005067

A:Map position: 14L

C:Keywords: transmembrane protein

F:598-614/Domain: transmembrane #status predicted <TMM>

Query Match 81.2%; Score 39; DB 2; Length 997;
Best Local Similarity 77.8%; Pred. No. 96;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
DB 232 SGGSSGSPV 240

RESULT 14
B82414
leucine aminopeptidase-related protein VCA0812 [imported] - Vibrio cholerae (strain NI69)

C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: B82414
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;

Chardson, D.; Ermolaeva, M.D.; Yamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
J. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: B82414
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-476 <HEI>
A:Cross-references: GB:AE004409; GB:AE003853; NID:g9658233; PIDN:AAF96710.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain NI6961; biotype El Tor

C:Genetics:
A:Gene: VCA0812
A:Map position: 2

Query Match 79.2%; Score 38; DB 2; Length 476;
Best Local Similarity 87.5%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGNSSGSPV 9
DB 316 GSSSGSPV 323

RESULT 15
T49812
hypothetical protein B11B22.140 [imported] - Neurospora crassa

C:Species: Neurospora crassa
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000

C:Accession: T49812
R:Schulte, U.; Aign, V.; Hohelsel, J.; Brandt, P.; Partmann, B.; Holland, R.; Nyakatu

submitted to the Protein Sequence Database, May 2000

A:Reference number: 225022
A:Accession: T49812

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-182 <SCH>
A:Cross-references: EMBL:AL356834; GSPDB:GN00116; NCSP:B11B22.140

A:Experimental source: BAC clone B11B22; strain OR74A

C:Genetics:
A:Gene: NCSP:B11B22.140

A:Map position: 6

A:Introns: 46/1; 80/1; 103/3; 140/3

Query Match 77.1%; Score 37; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGS 7
DB 161 TCGNSGS 167

Search completed: December 20, 2002, 12:11:31
Job time : 25.1579 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 05:38:27 ; Search time 13.7368 Seconds
(without alignments)
27.174 Million cell updates/sec

Title: US-10-008-355-26
Perfect score: 48
Sequence: 1 TCGNSGSPV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	100.0	336	1 STSP_STAUV	P04188 staphylococ
2	39	81.2	418	1 ER24_HUMAN	O76062 homo sapien
3	39	81.2	997	1 YNM3_YEAST	P53920 saccharomyc
4	37	77.1	316	1 GSEP_BACLI	P80057 bacillus il
5	37	77.1	401	1 HME1_MOUSE	P09065 mus musculu
6	36	75.0	320	1 CE11_AGABI	O00023 agarcus dl
7	36	75.0	378	1 PLTB_ASPNG	O00205 aspergillus
8	36	75.0	628	1 ABFA_ASPNG	P42254 aspergillus
9	35	72.9	108	1 Y108_NPVOP	O10347 orgyia pseu
10	35	72.9	417	1 IRIX5_HUMAN	P78411 homo sapien
11	35	72.9	521	1 YF92_MYCPN	O50335 mycoplasma
12	35	72.9	531	1 YF88_MYCPN	O50339 mycoplasma
13	35	72.9	584	1 YMB3_YEAST	O04228 saccharomyc
14	35	72.9	1714	1 STEP_DROME	O32870 mycobacteri
15	34	70.8	303	1 Y007_MYCLE	O97688 streptococc
16	34	70.8	325	1 YF65_STRPN	P29068 thermocactin
17	34	70.8	424	1 CBPT_THEVU	P09306 beet wester
18	34	70.8	607	1 V66K_BMYVF	O99211 homo sapien
19	34	70.8	843	1 AXN2_HUMAN	P09307 beet wester
20	34	70.8	1035	1 RRP0_BMYVF	P28507 saccharomyc
21	34	70.8	1286	1 RRV5_YEAST	O10571 homo sapien
22	34	70.8	1319	1 MNI_HUMAN	P00522 drosophila
23	34	70.8	1520	1 ABL_DROME	P28997 stronglylce
24	33	68.8	117	1 WNT6_STRPN	P75637 mycoplasma
25	33	68.8	221	1 YB31_MYCPN	P76633 escherichia
26	33	68.8	261	1 Y6CW_ECOLI	O00164 ictalurid h
27	33	68.8	281	1 VG11_HSV1	O07424 mus musculu
28	33	68.8	282	1 CDX4_MOUSE	P28009 drosophila
29	33	68.8	327	1 HMDL_DROME	P28994 agrocyphe ae
30	33	68.8	328	1 PYRD_DROME	O00374 colletoitic
31	33	68.8	366	1 OURA_CAUCR	P46239 fusarium ox
32	33	68.8	380	1 PLTA_COGLX	
33	33	68.8	385	1 GUNF_FUSOX	

34	33	68.8	511	1 MDJ1_YEAST	P35191 saccharomyc
35	33	68.8	529	1 SPI5_SPRGR	P19471 streptomyce
36	33	68.8	636	1 CA13_RAT	P13941 rattus norv
37	33	68.8	637	1 T4BA_BACCO	O07605 bacillus co
38	33	68.8	710	1 DA_DROME	P14420 drosophila
39	33	68.8	737	1 YD70_MYCPN	P75411 mycoplasma
40	33	68.8	848	1 CLPC_MYCLE	P24428 mycobacteri
41	33	68.8	848	1 CLPC_MYCTU	O06286 mycobacteri
42	33	68.8	921	1 CIA4_CANAL	O14427 candida alb
43	33	68.8	983	1 4ET_MOUSE	O9523 mus musculu
44	33	68.8	1007	1 Y741_CHLMU	O9916 chlamydia m
45	33	68.8	1014	1 NANH_CLOSE	P29767 clostridium

ALIGNMENTS

```

RESULT 1
ID STSP_STAUV STANDARD: PRT: 336 AA.
AC P04188:
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutamyl endopeptidase precursor (EC 3.4.21.19) (Staphylococcal serine
DE proteinase) (V8 proteinase) (Endoproteinase Glu-C).
GN SSFA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V8:
RX MEDLINE=87316953; PubMed=3306605;
RA Carmona C., Gray G.L.;
RT "Nucleotide sequence of the serine protease gene of Staphylococcus
RL aureus, strain V8."
RL Nucleic Acids Res. 15:6757-6757(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20569178; PubMed=11119502;
RA Rice K., Peralta R., Bast D., de Azavedo J., McGavin M.J.;
RT "Description of staphylococcus serine protease (ssp) operon in
RT Staphylococcus aureus and nonpolar inactivation of ssfa-encoded serine
RT protease."
RL Infect. Immun. 69:159-169(2001).
RN [3]
RP SEQUENCE OF 69-280.
RC STRAIN=V8:
RX MEDLINE=78212487; PubMed=96922;
RA Drapeau G.R.;
RT "The primary structure of staphylococcal protease."
RL Can. J. Biochem. 56:534-544(1978).
CC -!- FUNCTION: PREFERENTIALLY CLEAVES PEPTIDE BONDS ON THE CARBOXYL-
CC TERMINAL SIDE OF ASPARTATE AND GLUTAMATE.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Asp-I-Xaa, Glu-I-Xaa.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.
CC -!- DATABASE: NAME=worthington-enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/P/STAP.html".
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Y00356; CAA68434.1; -
CC EMBL: AF309515; ANG45843.1; -
CC FIC: A26812; PRSASK.
CC MEROPS: S01.269; -
CC InterPro: IPR000126; Ser__proteas_V8.

```



```

CC "The DNA sequence of cosmid 14-11b from chromosome XIV of
RT Saccharomyces cerevisiae reveals an unusually high number of
RL overlapping open reading frames."
CC -1- SIMILARITY: TO S.POMBE SPEC1685.05.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; Z69382; CAA93384.1; -.
DR EMBL; Z71399; CAA96004.1; -.
DR MEROPS; S01.OPC; -.
DR SGD; S0005067; YNL123W.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 1.
DR SMART; SM00228; PDZ; 2.
DR Hypothetical protein.
SQ SEQUENCE 997 AA; 110881 MW; A26005C1DDDB932C CRC64;
OY 1 TGGNGSPV 9 81.2% Score 39; DB 1; Length 997;
Db 232 SGGSSGSPV 240 77.8% Pred. NO. 25;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 4
GSEP_BACLI
ID GSEP_BACLI STANDARD: PRT; 316 AA.
AC P80057;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glutamyl endopeptidase precursor (EC 3.4.21.19) (Glutamate specific
DE endopeptidase) (GSE).
DE
DE
GN BLASE.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1402;
[1]
RN RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=ATCC 14580;
RC MEDLINE=93054737; PubMed=1429718;
RA Kakudo S., Kikuchi N., Kitadokoro K., Fujiwara T., Nakamura E.,
RA Okamoto H., Shin M., Tamaki M., Teroka H., Tszuzuki H., Yoshida N.;
RT "Purification, characterization, cloning, and expression of a
RT glutamic acid-specific protease from Bacillus licheniformis ATCC
RT 14580."
RT J. Biol. Chem. 267:23782-23788(1992).
RN RL
RN RP SEQUENCE OF 95-316.
RP MEDLINE=92155199; PubMed=1346764;
RA Svendsen I., Bredam K.;
RA "Isolation and amino acid sequence of a glutamic acid specific
RT endopeptidase from Bacillus licheniformis."
RL Eur. J. Biochem. 204:165-171(1992).
CC -1- FUNCTION: SPECIFIC FOR HYDROLYSIS OF PEPTIDES BONDS ON THE
CC CARBOXYL SIDE OF ACIDIC AMINO ACID RESIDUES, WITH A STRONG
CC PREFERENCE FOR GLU.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Asp-|-Xaa, Glu-|-Xaa.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```

```
CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10060; BAA00949.1; -.
DR PIR; S23078; S23078.
DR PIR; A45134; A45134.
DR MEROPS; S01.271; -.
DR InterPro; IPR000126; Ser_protease_V8.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin_1.
DR PRINTS; PR00839; VBPROTEASE.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 30 POTENTIAL.
FT PROPER 31 94
FT CHAIN 95 316 GLUTAMYL ENDOPEPTIDASE.
FT ACT_SITE 141 141 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 261 261 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 126 142
FT DISULFID 275 279
SQ SEQUENCE 316 AA; 33611 MW; 96D7552CB7089B09 CRC64;
OY 2 GGNSSGPV 9 77.1%; Score 37; DB 1; Length 316; Best Local Similarity 87.5%; Pred. No. 17; Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0.
Db 258 GGSGSPV 265
RN RESULT 5
HMEI_MOUSE STANDARD: PTR; 401 AA.
AC P09065;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Homeobox protein engrailed-1 (Mo-En-1).
GN EN1 OR EN-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93185339; PubMed=1363401; Logan C., Hanks M.C., Noble-Trophan S., Nallainathan D., Provatt N.J., Joyner A.L.;
RA "Cloning and sequence comparison of the mouse, human, and chicken engrailed genes reveal potential functional domains and regulatory regions.";
RT Dev. Genet. 13:345-358(1992).
RN [2]
RP SEQUENCE OF 278-401 FROM N.A.
RX MEDLINE=88112776; PubMed=2892757; Joyner A.L., Martin G.R.;
RA "En-1 and En-2, two mouse genes with sequence homology to the Drosophila engrailed gene: expression during embryogenesis.";
RL Genes Dev. 1:29-38(1987).
RN [3]
RP SEQUENCE OF 298-401 FROM N.A.
RX MEDLINE=86079501; PubMed=2416459; Joyner A.L., Kornberg T., Coleman K.G., Cox D.R., Martin G.R.;
RT "Expression during embryogenesis of a mouse gene with sequence homology to the Drosophila engrailed gene.";
RL Cell 43:29-37(1985).
RN [4]
```

```

RP SEQUENCE OF 321-380 FROM N.A.
RX MEDLINE=91099509; PubMed=1980115;
RA Holland P.W.H., Williams N.A.;
RT "Conservation of engrailed-like homeobox sequences during vertebrate
RL evolution.";
CC FEBS Lett. 277:250-252(1990).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE ENGRAILED HOMEBOX FAMILY.
CC -1- SIMILARITY: CONTAINS 1 HOMEBOX DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; L12703; AAA03660.2; -
DR EMBL; Y00201; CA68361.1; -
DR PIR; A26629; A26629.
DR PIR; A24778; A24778.
DR PIR; S13009; S13009.
DR PIR; A48423; A48423.
DR HSSP; P02836; 3HDD.
DR TRANSFAC; T02016; -
DR MGD; MGI:95389; En1.
DR InterPro: IPR000747; Engrailed.
DR InterPro: IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR PRINTS; PR00026; ENGRAILED.
DR PRINTS; PR00024; HOMEBOX.
DR ProDom; PD000010; Homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS50071; HOMEBOX_2; 1.
DR PROSITE; PS00033; ENGRAILED; 1.
DR Homeobox; DNA-binding; Developmental protein; Nuclear protein.
FT DOMAIN 52 87 PRO-RICH.
FT DOMAIN 73 87 POLY-PRO.
FT DOMAIN 207 228 POLY-ALA.
FT DNA_BIND 312 371 HOMEBOX.
SQ SEQUENCE 401 AA; 40950 MW; 1F90210950152FAE CRC64;

Query Match 77.1%; Score 37; DB 1; Length 401;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGCNGSP 8
DB 238 SGGNGSP 245

RESULT 6
CELL_AGABI STANDARD; PRT; 320 AA.
ID CELL_AGABI
AC Q00023;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cellulose-growth-specific protein precursor.
GN CELL.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RX MEDLINE=93012985; PubMed=1398098;
RA Raguz S., Yaguee E., Wood D.A., Thurston C.F.;
RT "Isolation and characterization of a cellulose-growth-specific gene
from Agaricus bisporus.";

```

```

RL Gene 119:183-190(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=D649;
RX MEDLINE=94237428; PubMed=8181702;
RA Arnesilla A.L., Thurston C.F., Yaguee E.;
RT "CELL: a novel cellulose binding protein secreted by Agaricus
RL bisporus during growth on crystalline cellulose.";
RL FEBS Microbiol. Lett. 116:293-299(1994).
CC -1- FUNCTION: PROBABLE GLYCOSYL HYDROLASE ACTIVE ON CELLULOSE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO FAMILY 61 OF GLYCOSYL HYDROLASES.
CC -1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; M86356; AA53434.1; -
DR HSSP; P00725; 2CBH.
DR InterPro: IPR000254; CBD_fungal.
DR InterPro: IPR005103; Glyco_hydro_61.
DR Pfam; PF00734; CBM_1; 1.
DR Pfam; PF03443; Glyco_hydro_61; 1.
DR SMART; SM00236; fCBD; 1.
DR PROSITE; PS00562; CBD_FUNGAL; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 320 CELLULOSE-GROWTH-SPECIFIC PROTEIN.
FT DOMAIN 30 261 CATALYTIC (POTENTIAL).
FT DOMAIN 262 285 LINKER (POTENTIAL).
FT DOMAIN 286 320 CELLULOSE-BINDING (POTENTIAL).
FT DISULFID 303 319 BY SIMILARITY.
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 320 AA; 33754 MW; 60E2C808089CA2B CRC64;

Query Match 75.0%; Score 36; DB 1; Length 320;
Best Local Similarity 75.0%; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCNGSP 8
DB 256 SGGNGSP 263

RESULT 7
PLYB_ASPNG STANDARD; PRT; 378 AA.
ID PLYB_ASPNG
AC Q00205;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Pectin lyase B precursor (EC 4.2.2.10) (PLB).
GN PELB.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 120.49 / M400;
RX MEDLINE=92357005; PubMed=1495474;
RA Kusters-Van Someren M., Filippi M., de Graaff L., van den Broeck H.,
RA Kester H., Hinnen A., Visser J.;
RT "Characterization of the Aspergillus niger pelb gene: structure and
RT regulation of expression.";
RL Mol. Gen. Genet. 234:113-120(1992).
CC -1- CATALYTIC ACTIVITY: Eliminative cleavage of pectin to give

```


DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Iroquois-class homeobox protein IRX-5 (Iroquois homeobox protein
 DE 5) (IRX-2a).
 GN IRX5 OR IRX2a.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99299262; PubMed=10370142;
 RA Lewis M.T., Ross S., Strickland P.A., Snyder C.J., Daniel C.W.;
 RT "Regulated expression patterns of IRX-2, an Iroquois-class homeobox
 RT gene, in the human breast."
 RL Cell Tissue Res. 296:549-554(1999).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE TALE/IRX HOMEBOX FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U90304; AAB50002.1; -;
 DR EMBL: U90309; AAB50007.1; -;
 DR HSSP: P41778; IDU6.
 DR TRANSFAC: T02436; -;
 DR GeneW: HGNC:14361; IRX5.
 DR MIM: 606195; -;
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR003893; Iroquois_homeo.
 DR Pfam: PF00046; homeobox_1.
 DR ProDom: PD00010; Homeobox_1.
 DR SMART: SM00389; HOX; 1.
 DR SMART: SM00548; IRO; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 KM Homeobox; DNA-binding; Nuclear protein.
 FT DNA_BIND 47 109 HOMEBOX (TALE-TYPE).
 FT DOMAIN 120 126 POLY-GLU.
 FT DOMAIN 254 260 POLY-PRO.
 FT DOMAIN 284 287 POLY-GLY.
 FT CONFLICT 145 145 G -> A (IN REF. 1; AAB50007).
 SQ SEQUENCE 417 AA: 43718 MW: 401F2D865C052AFA CRC64;
 Query Match 72.9%; Score 35; DB 1; Length 417;
 Best Local Similarity 85.7%; Pred. No. 51;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 GGNSSGP 8
 DB 286 GGNSSGP 292
 RESULT 11
 YF92_MYCPN STANDARD: PRT; 521 AA.
 AC 050335;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical lipoprotein MPN592 precursor (D02_ort521).
 GN MPN592 OR MP250.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Iroquois-class homeobox protein IRX-5 (Iroquois homeobox protein
 DE 5) (IRX-2a).
 GN IRX5 OR IRX2a.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99299262; PubMed=10370142;
 RA Lewis M.T., Ross S., Strickland P.A., Snyder C.J., Daniel C.W.;
 RT "Regulated expression patterns of IRX-2, an Iroquois-class homeobox
 RT gene, in the human breast."
 RL Cell Tissue Res. 296:549-554(1999).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: BELONGS TO THE TALE/IRX HOMEBOX FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U90304; AAB50002.1; -;
 DR EMBL: U90309; AAB50007.1; -;
 DR HSSP: P41778; IDU6.
 DR TRANSFAC: T02436; -;
 DR GeneW: HGNC:14361; IRX5.
 DR MIM: 606195; -;
 DR InterPro: IPR001356; Homeobox.
 DR InterPro: IPR003893; Iroquois_homeo.
 DR Pfam: PF00046; homeobox_1.
 DR ProDom: PD00010; Homeobox_1.
 DR SMART: SM00389; HOX; 1.
 DR SMART: SM00548; IRO; 1.
 DR PROSITE: PS00027; HOMEBOX_1; 1.
 DR PROSITE: PS00071; HOMEBOX_2; 1.
 KM Homeobox; DNA-binding; Nuclear protein.
 FT DNA_BIND 47 109 HOMEBOX (TALE-TYPE).
 FT DOMAIN 120 126 POLY-GLU.
 FT DOMAIN 254 260 POLY-PRO.
 FT DOMAIN 284 287 POLY-GLY.
 FT CONFLICT 145 145 G -> A (IN REF. 1; AAB50007).
 SQ SEQUENCE 417 AA: 43718 MW: 401F2D865C052AFA CRC64;

RX MEDLINE=96177562; PubMed=8604303;
 RA Hilbert H., Himmelfreich R., Piagens H., Herrmann R.;
 RT "Sequence analysis of 56 kb from the genome of the bacterium
 RT Mycoplasma pneumoniae comprising the dna region, the atp operon and a
 RT cluster of ribosomal protein genes."
 RL Nucleic Acids Res. 24:628-639(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;
 RA Himmelfreich R., Hilbert H., Piagens H., Pirkl E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
 RT pneumoniae."
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 CC (potentially).
 CC -1- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U43738; AAC43664.1; -;
 DR EMBL: AE000025; AAB95898.1; -;
 DR InterPro: IPR002414; DUF30731.
 DR Pfam: PF01727; DUF30; 1.
 DR Pfam: PF01732; DUF31; 1.
 DR PRINTS: PR00840; Y06768FAMILY.
 KM Hypothetical protein; Lipoprotein; Membrane; Signal;
 KM Complete proteome.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 521 HYPOTHETICAL LIPOPROTEIN MPN592.
 FT LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
 SQ SEQUENCE 521 AA: 59500 MW: 0E706DC8EC8EBDA CRC64;
 Query Match 72.9%; Score 35; DB 1; Length 521;
 Best Local Similarity 75.0%; Pred. No. 65;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 GGNSSGPV 9
 DB 423 GGNSSGSPV 430
 RESULT 12
 YF88_MYCPN STANDARD: PRT; 531 AA.
 AC 050339;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical lipoprotein MPN588 precursor (D02_ort531).
 GN MPN588 OR MP254.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE=96177562; PubMed=8604303;
 RA Hilbert H., Himmelfreich R., Piagens H., Herrmann R.;
 RT "Sequence analysis of 56 kb from the genome of the bacterium
 RT Mycoplasma pneumoniae comprising the dna region, the atp operon and a
 RT cluster of ribosomal protein genes."
 RL Nucleic Acids Res. 24:628-639(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;


```

RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -----
CC -1- SIMILARITY: BELONGS TO THE MG067 / MG068 / MG395 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 043738; AAC43668.1; -
CC EMBL: AE000025; AAB95902.1; -
CC InterPro: IPR002414; DUF30/31.
CC Pfam: PF01727; DUF30; 1.
CC DR Pfam: PF01732; DUF31; 1.
CC PRINTS: PR00840; Y06768FAMILY.
CC Hypothetical protein; Lipoprotein; Membrane; Signal;
CC Complete proteome.
CC SIGNAL 1 22 POTENTIAL.
CC CHAIN 23 531 HYPOTHETICAL LIPOPROTEIN MPN588.
CC LIPID 23 23 N-ACYL DIGLYCERIDE (POTENTIAL).
CC SEQUENCE 531 AA: 60141 MW: 0A3A089261B3C998 CRC64;
SQ
Query Match 72.9%; Score 35; DB 1; Length 531;
Best Local Similarity 75.0%; Pred. No. 67;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 GGNSSGPV 9
DB 424 GGSSEPL 431

```

```

SQ SEQUENCE 584 AA: 66761 MW: DA4DD3D624C17B8 CRC64;
Query Match 72.9%; Score 35; DB 1; Length 584;
Best Local Similarity 85.7%; Pred. No. 74;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 GGNSSGP 8
DB 172 GGNSSGP 178

```

RESULT 14
SYEP_DROME
ID SYEP_DROME STANDARD: PRT: 1714 AA.
AC P28668; O9YCF5;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bifunctional aminoacyl-tRNA synthetase [includes: Glutamyl-tRNA
DE synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase); Prolyl-tRNA
DE synthetase (EC 6.1.1.15) (Proline--tRNA ligase)].
GN AATS-GLUPRO OR CG5394.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92097547; PubMed=1756734;
RA Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M.;
RT "A component of the multisynthetase complex is a multifunctional
RT aminoacyl-tRNA synthetase";
RL EMBO J. 10:4267-4277(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Oregon-R;
RX MEDLINE=97217441; PubMed=9063462;
RA Cerini C., Semeriva M., Gratecos D.;
RT "Evolution of the aminoacyl-tRNA synthetase family and the
RT organization of the Drosophila glutamyl-prolyl-tRNA synthetase gene.
RT Introns/exon structure of the gene, control of expression of the two
RT mRNAs, selective advantage of the multienzyme complex.";
RL Eur. J. Biochem. 244:176-185(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blaise R.G., Chapple M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bonshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaisli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sieden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Massaman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*.";
Science 287:2185-2195(2000).
CC -i- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(glu) = AMP +
CC dihydrophosphate + L-glutamyl-tRNA(glu).
CC -i- CATALYTIC ACTIVITY: ATP + L-proline + tRNA(pro) = AMP +
CC dihydrophosphate + L-prolyl-tRNA(pro).
CC -i- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS
CC COMPRSED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE
CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,
CC ARGINYL, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY
CC PROTEINS, P18, P48 AND P43.
CC -i- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -i- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II
CC AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -i- SIMILARITY: CONTAINS 6 WHEP-TRS DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M74104; AAA28594.1; -;
DR EMBL: U59923; AAC47469.1; -;
DR EMBL: AE003745; AAF56211.1; -;
DR PIR: S18644; S18644.
DR HSSP: P00962; 1GTR.
DR FlyBase: FBgn0005674; Aats-glupro.
DR InterPro: IPR002106; AACRNA_ligaseII.
DR InterPro: IPR004046; GST_Cterm.
DR InterPro: IPR004526; GLX_arch.
DR InterPro: IPR000924; GLX_arch.
DR InterPro: IPR004154; HGTP_antlicodon.
DR InterPro: IPR004499; PROS_fam_I.
DR InterPro: IPR000738; WHEP-TRS.
DR InterPro: IPR002314; tRNA-synt_2b.
DR InterPro: IPR001412; tRNA-synt_1.
DR InterPro: IPR002316; tRNA-synt_pro.
DR Pfam: PF00043; GST_C; 1.
DR Pfam: PF00458; WHEP-TRS; 6.
DR Pfam: PF00587; tRNA-synt_2b; 1.
DR Pfam: PF00749; tRNA-synt_1c; 1.
DR Pfam: PF03129; HGTP_antlicodon; 1.
DR PRINTS: PR00987; TRNASYNTHGLU.
DR PRINTS: PR01046; TRNASYNTHPRO.
DR TIGRfams: TIGR00408; pros_fam_I; 1.
DR TIGRfams: TIGR00463; glx_arch; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
DR PROSITE: PS50862; AA_TRNA_LIGASE_II; 1.
DR PROSITE: PS00762; WHEP-TRS; 6.
KW Aminoacyl-tRNA synthetase; protein biosynthesis; ligase; ATP-binding;
KW Multifunctional enzyme; Repeat.
KW DOMAIN 170 754 GLUTAMYL-TRNA SYNTHETASE.
FT DOMAIN 755 800 WHEP-TRS 1.
FT DOMAIN 827 872 WHEP-TRS 2.
FT DOMAIN 901 946 WHEP-TRS 3.
FT DOMAIN 980 1025 WHEP-TRS 4.
FT DOMAIN 1055 1100 WHEP-TRS 5.
FT DOMAIN 1129 1173 WHEP-TRS 6.

FT DOMAIN 1174 1180 POLY-GLY.
FT DOMAIN 1207 1214 PROLYL-TRNA SYNTHETASE.
FT SITE 209 220 "HIGH" REGION.
FT SITE 438 442 "KMSKS" REGION.
FT BINDING 441 441 ATP (BY SIMILARITY).
FT CONFLICT 102 106 TSPLE -> DRSIA (IN REF. 3).
FT CONFLICT 233 234 VC -> AP (IN REF. 3).
FT CONFLICT 341 345 NTACA -> KCVLR (IN REF. 3).
FT CONFLICT 583 583 K -> R (IN REF. 3).
FT CONFLICT 583 583 L -> S (IN REF. 3).
FT CONFLICT 692 692 T -> S (IN REF. 3).
FT CONFLICT 753 753 T -> S (IN REF. 3).
FT CONFLICT 802 802 P -> S (IN REF. 3).
FT CONFLICT 873 873 P -> T (IN REF. 3).
FT CONFLICT 887 887 G -> V (IN REF. 3).
FT CONFLICT 1201 1201 P -> PA (IN REF. 3).
FT CONFLICT 1461 1461 MISSING (IN REF. 3).
FT CONFLICT 1587 1587 G -> V (IN REF. 3).
SQ SEQUENCE 1714 AA; 189197 MW; 6F8C58045E4B8C CRC64;

Oy 1 TCGNSGSPV 9
Db 1037 TCGDSGSDV 1045

Query Match 72.9%; Score 35; DB 1; Length 1714;
Best Local Similarity 77.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 15
Y007_MYCLE
ID Y007_MYCLE STANDARD; PRT; 303 AA.
AC 032870;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ML0007.
GN ML0007 OR MLB1770.07.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN.
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -i- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV0007.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Z70722; CA94717.1; -;
DR EMBL: AL583917; CAC29515.1; -;
DR Leproma; ML0007; -;

KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 205 225 POTENTIAL.
 FT TRANSMEM 264 284 POTENTIAL.
 SQ SEQUENCE 303 AA; 32236 MW; 9761DD87391C1E4 CRC64;

Query Match 70.8%; Score 34; DB 1; Length 303;
 Best Local Similarity 77.8%; Pred. No. 55;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGGNGSPV 9
 Db 246 TGGNGSELY 254

Search completed: December 20, 2002, 12:02:24
 Job time : 16.7368 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 17:28:51 ; Search time 140 Seconds

(without alignments)
3072.369 Million cell updates/sec

Title: US-10-008-355-2_COPY-522_712

Perfect score: 1003

Sequence: 1 SKSVIAARAIAQADAMANAY.....LFMIDKQGCPRLLIQELKLI 191

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPFO/spool/US10008355/runat_17122002_112703_16900/app.query.fasta_1.327
-DB=N.Geneseq_101002 -QFWT=fastap -SUFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -STAR=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pcr -NORF=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10008355 @CGN.1.1.0 @runat_17122002_112703_16900 -MCP=6 -ICPU=3
-NO_XIPXY -NO_MMAR -LARGESQUEY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: N.Geneseq_101002.*
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description
1	1003	100.0	2139	24	AA143635	Porphyromonas ging
2	90	9.0	2136	17	AA11244	Neisseria meningit
3	89.5	8.9	2794	22	AAH17774	Human cDNA sequenc
4	89.5	8.9	3259	22	AA156418	Human poly nucleoti
5	89.5	8.9	3514	22	AAH16079	Human cDNA sequenc
6	89.5	8.9	3899	23	AAH86370	DNA encoding novel
7	89.5	8.9	3915	22	AAH61204	Human poly nucleoti
8	89.5	8.9	6258	23	AAH1791	DNA encoding novel
9	89.5	8.9	10862	19	AAH34396	Yellow fever infec
10	89.5	8.9	10892	22	AAH83821	Chimeric DNA from
11	86	8.6	2365589	24	ABA30521	Genomic sequence o
12	85.5	8.5	651	24	ABQ65948	Arabidopsis thalia
13	85.5	8.5	972	21	AAH46293	Arabidopsis thalia
14	83	8.3	606	22	AAH53441	S. epidermidis thali
15	83	8.3	731	22	AAH91391	N. meningitidis ope
16	83	8.3	731	24	ABK37771	DNA sequence upstr
17	83	8.3	1083	24	ABN92057	Staphylococcus epi
18	83	8.3	3003	22	AAH54900	S. epidermidis gen
19	82.5	8.2	319630	24	ABQ67194	Listeria innocua c
20	82.5	8.2	3253	21	AAH75951	Human OREF ORF1506
21	81.5	8.1	720	23	AAH51714	Staphylococcus aur
22	81.5	8.1	765	23	AAH54542	Staphylococcus aur
23	81.5	8.1	927	24	ABN65304	Streptococcus poly
24	81.5	8.1	2155561	24	ABN71527	Streptococcus poly
25	81	8.1	849	21	AAH01293	Ehrlichia canis im
26	81	8.1	849	24	ABK68853	DNA encoding Ehrli
27	81	8.1	2037	21	AAH01295	Ehrlichia canis im
28	80.5	8.0	3951	23	AAH82910	DNA encoding novel
29	80.5	8.0	14105	23	ABH28585	Drosophila melanog
30	80.5	8.0	17769	23	ABH28585	Drosophila melanog
31	80	8.0	1019	18	AAH75063	Staphylococcus aur
32	80	8.0	1558	13	AAO27988	Protease from S. A
33	80	8.0	1586	13	AAO27988	Protease from S. A
34	79.5	7.9	861	24	ABK75469	Bacillus lichenifo
35	79.5	7.9	2244	20	AAH93649	Nucleic acid seque
36	79.5	7.9	2365589	24	ABA30521	Genomic sequence o
37	79	7.9	1833	18	AAH73363	Konjak mosaic viru
38	78.5	7.8	1497	21	AAH54262	Neisseria meningit
39	78.5	7.8	6444	18	AAH74544	Staphylococcus aur
40	78	7.8	2017	22	AAH99629	Human protein enco
41	78	7.8	2541	21	AAH42943	Arabidopsis thalia
42	78	7.8	2643	22	AAH94354	Human full-length
43	78	7.8	3192	22	AAH16193	Human cDNA sequenc
44	78	7.8	10976	21	ABH50890	Japanese encephali
45	77.5	7.7	905	23	AAH80158	DNA encoding novel

ALIGNMENTS

RESULT 1
AA143635
AA143635 standard; DNA; 2139 BP.

AA143635;
05-SEP-2002 (first entry)

Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) coding sequence.

Dipeptidylpeptidase-7; DPP-7; gene; ds; enzyme; amidolytic cleavage;

DPP-7 inhibitor identification; periodontal disease; gingivitis;

periodontitis.

Porphyromonas gingivalis.
Key Location/Qualifiers
CDS 1..2139

FT		/tag= a
FN		/product= "Porphyromonas gingivialis DPP-7"
XX		
PD	W0200238742-A2.	
XX		
PD	16-MAY-2002.	
XX		
PF	08-NOV-2001; 2001MO-US46782.	
XX		
FR	08-NOV-2000; 2000US-246827P.	
XX		
PA	(UYGE-) UNIV GEORGIA RES FOUND INC.	
PI	Travis J, Potempa JS, Banbula A, Bugno M;	
DR	WPI; 2002-490075/52.	
XX	P-P5DB: AAOI5205.	
PT	Novel isolated dipeptidylpeptidase useful for identifying inhibitor of	
PT	the dipeptidylpeptidase for protecting an animal from periodontal	
PS	disease caused by Porphyromonas gingivialis -	
XX	ClaIm 11; Fig 4; 65pp; English.	
CC	The invention comprises the amino acid and coding sequence of the	
CC	Porphyromonas gingivialis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7	
CC	enzyme has amidolytic activity for cleavage of a peptide bond between the	
CC	second and third amino acids from the N'-terminal end of a target peptide.	
CC	The DPP-7 target peptide has an aliphatic or aromatic residue as a	
CC	substituent on the alpha-carbon atom of the second amino acid from the	
CC	N-terminal end. The DPP-1 protein and DNA sequences of the invention are	
CC	useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for	
CC	reducing the growth of a bacterium and protecting an animal from a	
CC	periodontal disease caused by Porphyromonas gingivialis (e.g., gingivitis	
CC	or periodontitis). The present DNA sequence encodes the Porphyromonas	
CC	gingivialis dipeptidylpeptidase-7 (DPP-7) enzyme of the invention.	
XX		
SQ	Sequence 2139 BP; 544 A; 543 C; 565 G; 487 T; 0 other:	
Alignment Scores:		
Pred. No.:	2.04e-113	Length: 2139
Score:	1003.00	Matches: 191
Percent Similarity:	100.00%	Conservative: 0
Best local Similarity:	100.00%	Mismatches: 0
Query Match:	100.00%	Indels: 0
DB:	24	Gaps: 0
US-10-008-355-2.COPY_522_712 (1-191) x AAU43635 (1-2139)		
OY	1 SerTSSerValIleAlaIleAlaIleGlnAlaAspAlaMetAlaAsnLafTr 20	
Db	1564 TTCAGAGCGTAATAGCTGTCTGCCGCTATTCAAGCCGATGGCCAAATGCTCAT 1623	
OY	21 AlaIleIubLySgLYsArGLEuPhenheAagLYLeuArgGLuMeTYrProGLyArg 40	
Db	1624 GCCATTTAGAAGGGCAAGCCTCTTTCTTGCCGGTTTCGGAGATGTACCCCGAGCT 1683	
OY	41 AlaLeuProSerAspAlaAsnPhetThrMetArgMetSerTYrGLYSerTlyLeuSLyTYr 60	
Db	1684 GCTTGGCGGAGCGATGCCACTTCACCAAGCGATGAGCTACGCCTCATCAAGGATAT 1743	
OY	61 GluProGlnAspGLyaLatrPYrAsnTYrHISThrThGLyLSGLValLenGLuys 80	
Db	1744 GAACCGAGAGCGGTGCTGTGTACAATTCATACACAGGGAAGGGCGATTGGAGAG 1803	
OY	81 GlnAspProLySserAspGluPhenAlaValALngLnAsnLIleLeuAspLeuPhenArgThr 100	
Db	1804 CAGATCTCTAAGAGCGATGTGTTGGCGTACAGGAATAATCTCTGACCTCTTCCGAC 1863	
OY	101 LysAsnTYrGLyArGLyrrAlaGLuAsnGLyGlnLeuHisILAlaPhenLeuSerAsn 120	
Db	1864 AAAAACATATGATGGCTATGGCGGAACGGGTGACCTCCATATGCTTCTATCGAAMAC 1923	

QY	121	AspIlehrGrlgYlgYsAnSerGlySerProValPheAspSLyGAsNGlYARgIleuIleGly	140		
Db	1924	GACATTCACGGGGCGTTAACTCCGGTAGCCCGTATGTGATTAAGACGCGCTGTGATCGGT	1983		
QY	141	LeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeu	160		
Db	1984	CTTCCTTTCATGTGCAACTGGAACTATGAGTGGTGCATCGAGTTCGAACCCGATCTG	2043		
QY	161	GlnArGThrIleSerValAspIleArGTYrValLeuPheMetIleAspLYstrpGlyGln	180		
Db	2044	CAGCGCACAAATCAGCGCGTGCATCCGCTACGTTCTCTCATGATTGACAAATGCGGCTAG	2103		
QY	181	CysProArGleuIleGlnGluMetLysIleuIle	191		
Db	2104	TGCCCCGCTCATCCAAAGAGCTGAAGTTGATC	2136		
RESULT 2					
ID	AA711244				
XX	AA711244	standard; DNA: 2136 BP.			
AC	AA711244:				
XX	02-SEP-1996	(first entry)			
DT	Neisseria meningitidis strain B2163 transferrin receptor Tbp2 gene.				
XX					
DE	Transferrin receptor; Tbp2 subunit; deletion mutant; vaccine;				
XX	passive immunisation; immunotherapy; IM2169; IM2394; ss.				
XX	Neisseria meningitidis (strain B2163).				
OS					
XX					
FH	Key	Location/Qualifiers			
FT	CDS	1..2136			
FT		/*tag= a			
FT	sig_peptide	1..60			
FT		/*tag= b			
FT	mat_peptide	61..2133			
FT		/*tag= c			
XX					
PN	MO9533049-A2.				
XX					
PD	07-DEC-1995.				
XX					
PF	30-MAY-1995;	95WO-FR00701.			
XX					
PR	31-MAY-1994;	94FR-0006594.			
XX					
PA	(INMR) PASTEUR MERIEUX SERUMS & VACCINS.				
PA	(TRGE) TRANSGENE SA.				
P1	Jacobs E, Legrain M, Mazarin V, Lissole L;				
P1	Miller MBJ;				
XX					
DR	WPI: 1996-030562/03.				
XX					
DR	P-PSDB: AAR8649.				
XX					
PT	Polypeptide(s) for vaccination against Neisseria meningitidis group				
PT	B - comprising deletion mutants of transferrin receptor Tbp2				
PT	subunit				
XX					
PS	Disclosure; Page 82-87; 11app; French.				
XX					
CC	The present sequence is that of the N.meningitidis strain B2163				
CC	transferrin Tbp2 subunit gene. The Tbp2 polypeptide has three				
CC	domains: an N-terminal domain, a hinge domain and a C-terminal				
CC	domain; deletion mutants in which at least one of the domains is				
CC	partially or totally deleted are claimed, provided that the first				
CC	and second domains are not simultaneously partially or totally				
CC	deleted. The positions of the 3 domains in B2163 are defined by				
CC	alignment with the IM2169 sequence. The deletion mutant polypeptides				
CC	of the invention can generate an immune response against				
XX	N.meningitidis.				

QY	125	---GlyAsnSerCylserProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe	143
Db	778	TGTGGAAGACGTGGAACCTGCTTTTACCAGAAATGGAAGGATGGCATTTGCTTTC	837
QY	144	AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr	163
Db	838	CGAATCCAGAGGAACCTTGGGGGCTCAGGCCCTCATCTCATGTCTGCTGCAACAAT	897
QY	164	IleSerValAspIleArgTyr	170
Db	898	TGCGACGTGAGTCAACTTC	918
RESULT 4			
AA159418	ID	AA159418	standard; cDNA; 3259 BP.
XX	XX	AA159418;	
AC	XX	22-OCT-2001	(first entry)
DT	XX	Human polynucleotide SEQ ID NO 1621.	
DE	XX	Human; nootropic; immunosuppressant; cytoslatic; gene therapy; cancer;	
KM	XX	peripheral nervous system; neuropathy; central nervous system; CNS;	
KM	XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KM	XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KM	XX	chemokinetiic; thrombolytic; drug screening; arthritis; inflammation;	
KM	XX	leukaemia; ss.	
OS	XX	Homo sapiens.	
XX	XX	WO200153312-A1.	
PN	XX	26-JUL-2001.	
PD	XX	26-DEC-2000; 2000WO-US34263.	
PF	XX	21-JAN-2000; 2000US-0488725.	
PR	XX	25-APR-2000; 2000US-0552317.	
PR	XX	09-JUL-2000; 2000US-0598042.	
PR	XX	19-JUL-2000; 2000US-0620312.	
PR	XX	03-AUG-2000; 2000US-0653450.	
PR	XX	14-SEP-2000; 2000US-0662191.	
PR	XX	19-OCT-2000; 2000US-0693036.	
PR	XX	29-NOV-2000; 2000US-0727344.	
PA	XX	(HSE-) HXSEQ INC.	
PI	XX	Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	XX	Zhao QA, Zhou P, Goodrich R, Dermanic RT;	
DR	XX	WPT: 2001-442253/47.	
DR	XX	P-PSDB: AAM40262.	
PT	XX	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	XX	such as central nervous system injuries -	
PS	XX	Claim 1: SEQ ID NO 1621: 10078pp; English.	
XX	XX	The invention relates to human nucleic acids (AA157798-AA161369) and	
CC	XX	the encoded polypeptides (AAM38642-AAM42213) with nootropic,	
CC	XX	immunosuppressant and cytoslatic activity. The polynucleotides are useful	
CC	XX	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	XX	of the invention may be used to treat diseases of the peripheral nervous	
CC	XX	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	XX	localised neuropathies and central nervous system diseases, such as	
CC	XX	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	XX	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	XX	utilisation of the activities such as: Immune system suppression,	
CC	XX	Activin/Inhibin activity, chemotactic/chemokinetiic activity, haemostatic	
CC	XX	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	

CC	assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed specification.
CC	
xx	
SQ	Sequence 3259 BP; 761 A; 999 C; 898 G; 561 T; 0 other;
	Alignment Scores:
	Pred. No.: 1.5 Length: 3259
	Score: 89.50 Matches: 42
	Percent Similarity: 38.32% Conservative: 22
	Best Local Similarity: 25.15% Mismatches: 56
	Query Match: 8.92% Indels: 47
	DB: 22 Gaps: 8
US-008-355-2_COPY_522_712 (1-191) x AA159418 (1-3259)	
Oy	19 AATATrAlaIleGluLySGLyArGLeuPheAlaGLyLeuNrgLumEtyrPro 38
Db	918 ACCTATGGGGGTACAGAAAGGGCGGTGTACTTCGTGCAGATCAAGTCAATCAGAAAATCTCC 977
Oy	39 GLAArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52
Db	978 GTGAAGCACCTTCCTCCGTACAGAACCTCGACCACCTGCTGTCGTATCGCGTGCTCCTG 1037
Oy	53 ---SerTYrGLySerILeLYSGlyTYrGLuPrOGlnAspLyAlaTrPTyrASnTYrHIs 71
Db	1038 GACTCTCTCAGACACCCAGCTAGCGGAGAGACCTTTCTCC-----TATGGCTATGGA 1088
Oy	72 ThrTPrGLyLSGLYValLeuGLuLYSGlnAspProLYSerASPoLUphelAlaVAlGIn 91
Db	1089 GGCACTGGGAG-----AATGCCACCAAT----- 1112
Oy	92 GluAsnILeUAsPLeuPheArGTThrLYSAnTYrGLy---ArgTYrAlaGLuAsnGLY 110
Db	1113 -----AGCGGCTTTGAATACTACGGAGACAGACTTTCAGAG----- 1148
Oy	111 GlnLeuHISILeAlaPheLeuSERasnaSnAspILeThrGLY----- 124
Db	1149 -----AACGATGTATTGGCTGCTTTGCGGATTTTGAA 1181
Oy	125 ---GlyAsnSerGLySerProVALpheAspLYSAsnGLYArGLeuILleGLyLeuAlaPHe 143
Db	1182 TGCTGGAAATGACGTGGAACGTCTTTTACCAAAGAATGGAAGAGTGGCATTTCTTCC 1241
Oy	144 AspGLyaSnTrpGLuaLaMeTSerGLYAspILeGLUPheGLuPrOAspLeuGLInArGTThr 163
Db	1242 CGAATCCGAAGGAAGACCTTGGGGGCTCAGGCCCTCTATCTCATATCTCCTGGTGAAGAAT 1301
Oy	164 IleserValaSpILearGTyr 170
Db	1302 TGCCGACGTGAGATTCAACTTC 1322
RESULT 5	
ID	AAHL6079 standard; cDNA; 3514 BP.
XX	
AC	AAHL6079;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA sequence SEQ ID NO:14778.
XX	
KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	

CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 3899 BP; 798 A; 1052 C; 1108 G; 941 T; 0 other;

Alignment Scores:	
Pred. No.:	1, 92
Score:	89.50
Percent Similarity:	38.32%
Best Local Similarity:	25.15%
Query Match:	8, 92%
DB:	23
Gaps:	8
Length:	389
Matches:	42
Conservative:	22
Mismatches:	56
Indels:	47

US-10-008-355-2_COPY_522_712 (1-191) x AAS86370 (1-3899)

Qy	19	AllyrAlaIleGluIuysGlyLysAsnGleuPhePhaIleGlyLeuArGluMetYrPro	38
Db	2982	AGCTATGGGGGACAGAAAGGGGCGGTGATCTCTTCAGATGACATGAATGAAGAAATCTCC	2923
Qy	39	GlyArGAlaLeuProSerAspAla-----AsnPheThrMetArgMet-----	52
Db	2922	GTTGAAGACACCTTCGCTGTACAGAGCCGTGAACCCACAGTGGTCCGATGGCGTCCCTG	2863
Qy	53	---SerTyrGleSerIleLysGlyTYrGluProGlnAspGlyAlaTrpTYrAsnTYrHis	71
Db	2862	GACCTCCGACAGCCAGCCAGCTAGGAGCAAGACCTTCTCC-----TATGGCATATGGA	2812
Qy	72	ThrThrGlyLysGlyValLeuGluIuysGlnAspProLysSerAspGluPheAlaValGln	91
Db	2811	GGCAGCTGGGAAG-----AACTCCACCAAT-----	2788
Qy	92	GluAsnIleLeuAspLeuPheArGThrLysAsnTYrGly---ArgTYrAlaGluAsnGly	110
Db	2787	-----AACCGGTTTGAAAACTTCAAGACAGAGTTGCAGAG-----	2752
Qy	111	GlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThGly-----	124
Db	2751	-----AACGATGGTATTTGGCTTGGCGATTTTGAA	2719
Qy	125	---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe	143
Db	2718	TGTTGGAATGACGTGCAACTGCTTTTACCAGAAAGAAAGATGGGCACTTCTTTC	2659
Qy	144	AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr	163
Db	2658	CGATTCAGAAAGAGACGCTTGGGGGGTCAGGCCCTCATCTCATCTCTGTGTGAAGAT	2599
Qy	164	IleSerValAspIleArgTYr	170
Db	2598	TGCGCACTGGAGTTCAACTTC	2578
RESULT 7			
AA161204/c			
ID	AA161204	standard; cDNA; 3915 BP.	
XX	AA161204:		
AC	XX		
XX	XX		
XX	22-Oct-2001	(first entry)	
DE	Human	polynucleotide SEQ ID NO 5193.	
XX			
KW	Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia; ss.		
XX			
OS	Homo sapiens.		
XX			
XX	MO200153312-A1.		
XX			

XX 26-JUL-2001.
ED
XX
XX
PF 26-DEC-2000; 2000MO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX
PA (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
DR P-PSDB; AAM42048.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Claim 1; SEQ ID NO 5193; 10078bp; English.

CC The invention relates to human nucleic acids (AA157799-AA161369) and
CC the encoded polypeptides (AA136642-AA442213) with neurotrophic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

Sequence 3915 BP; 798 A; 1041 C; 1122 G; 954 T; 0 other;

Alignment Scores:	
pred. NO.:	1 93
Score:	89.50
Percent Similarity:	38.32%
Best Local Similarity:	25.15%
Query Match:	8.92%
DB:	22
Gaps:	8
Length:	3915
Matches:	42
Conservative:	22
Mismatch:	56
Indels:	47
Gaps:	8

US-10-008-355-2_COPY_522_712 (1-191) x AAI61204 (1-3915)

[illegible]

[illegible]

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0152370.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

```
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Alignment Scores:
Pred. No.: 0.899 Length: 972
Score: 85.50 Matches: 42
Percent Similarity: 38.16% Conservative: 16
Best Local Similarity: 27.63% Mismatches: 67
Query Match: 8.52% Indels: 27
DB: 21 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x AAC46293 (1-972)

QY 45 AspAlaAsnPhetHrMetArqMetSerTYrGlySerIleLysGlyTYrGluPro---Gln 63
DB 502 GATGCTAAGGAGCAAGATTATACAGAGGAAATTTGGTCTGTGATCCAGATAAT 561
QY 64 AsnGlyAlaIleTyrAsnTYrHsThrHnGlyLys-----Gly 76
DB 562 GATCTAGCTGTGTGTTGAAGATTGAAGGAGCGTAACCTAATCCTGTTCTCGGT 621
QY 77 ValLeuGluLysGlnAspProLysSerAspGluPheAlaVal-----Gln 91
DB 622 ACCTCCATGATGTACGCGTAGGTCAAGTGTCTTGGCATTTGGATATCATATGATAT 681
QY 92 GluAsnIleLeuAspLeuPheArqHrLysAsnTYrGlyArqTYr-----AlaGluAsn 109
DB 682 GAAAGACCTTGACATAGAGGGTAGTAGTGGGTTGGAAAGAGATACCTTCACCTAAT 741
QY 110 GlyGln---LeuHsIleAlaPheLeuSerAsnAsnAspIleHrGlyLysAsnSerGly 128
DB 742 GGGAAAGATATTTCACAGACCTATACAAACCGATGCTGATATTATACAGCAATTTGG 801
QY 129 SerProValPheAspLysAsnGlyArqLeuIleGlyLys-----AlaPheAspGly 145
DB 802 GGGCCATGTGCTGATCTTATGGCCATACCATAGGTGTGAACACGTCACATTCACCCGA 861
QY 146 AsnTyrGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArqHrIleSer 165
DB 862 AAGGAGGTGTATGTCTTCTGCTGTTACTTT-----GCCATTTCC 903
QY 166 ValAsp-----IleArqTYrValLeuPheMetIle 175
DB 904 ATTGACACAGATTGTCCGACACAGTTCCGTACCTCATTT 939

RESULT 14
AAH53441
ID AAH53441 standard; DNA; 606 BP.
XX
AC AAH53441:
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2275.
XX
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX vaccination; endocarditis; ds.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
```

```
XX
PF 09-NOV-2000; 2000MO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAXO ) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
DR WPI: 2001-316495/33.
DR P-PSDB: AAG82591.
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
PS Claim 8; Page 605-606; 2188pp; English.
XX
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX CC AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX CC N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 606 BP; 209 A; 92 C; 129 G; 176 T; 0 other;

Alignment Scores:
Pred. No.: 0.961 Length: 606
Score: 83.00 Matches: 34
Percent Similarity: 35.25% Conservative: 15
Best Local Similarity: 24.46% Mismatches: 46
Query Match: 8.28% Indels: 44
DB: 22 Gaps: 6

US-10-008-355-2_COPY_522_712 (1-191) x AAH53441 (1-606)

QY 57 IleLysGlyTYrGluProGlnAspGlyAlaTyrTYrAsnTYrHsThrHnGlyLysGly 76
DB 1 ATGAAAGGGGGCCCATCCATGAAAAACATCAAGAAACCTTGATTTTAAAGTAAGTCA 60
QY 77 ValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 96
DB 61 TTGCTAAAGAGTAGATGATCTTACAGGTGAAGATT-----GAGGCTCTAATGCAT 111
QY 97 LeuPheArqHr---LysAsnTYr-----GlyArqTYrAlaGluAsn 109
DB 112 TTTGCTATGACATTAATAAAATATTAACAACAGCCACACCATCGATATTATAGAGGT 171
QY 110 GlyGlnLeu----- 112
DB 172 AAGAAATTTGCTTTACTCTTCGAAAAAGACATCTACTCGAGCCGTCCGCAATTACAGTT 231
QY 113 -----HisIleAlaPheLeuSerAsnAsnAspIleHrGlyLys 125
DB 232 GCATCTATTGATCTAGTGCACACCTGAATTTTAAAGGAAAAATGATTCATTAATAGCA 291
QY 126 AsnSerGlySerProValPheAspLysAsnGlyArqLeuIleGlyLysAlaPheAspGly 145
DB 292 AAAAAAGATCT-----CTTGAGGATACTGCTAAAGTTTAGCGAGAAATCTTTGATGCA 345
```


Db 346 -----ATTGATTTTAAAGGTTTTCCCAAAAACCTGT 378
Qy 146 AsnTrpGlnIawetSerLysAspIleGluProAspLeuArgThrIle 164
 |||||
ID AAF91391 standard; DNA; 731 BP.
XX AAF91391.
AC AAF91391.
DT 04-MAY-2001 (first entry)
XX N. meningitidis (serogroup B) TbpA gene upstream sequence, SEQ ID:17.
DE Modified Gram-negative bacterium; outer membrane vesicle; bleb; vaccine;
KW genetically modified; protective antigen expression; LPS detoxification;
KM LPS; Lipid A; homologous recombination vector; immunisation;
KW immunoprotective; non-toxic; paediatric; cyclic; circular; ds.
XX Neisseria meningitidis.
OS Neisseria meningitidis.
PN WO200109350-A2.
XX 08-FEB-2001.
PD 31-JUL-2000; 2000WO-EP07424.
PF 03-AUG-1999; 99GB-001B319.
PR (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
PA Berthel FJ, Dalemans WILJ, Denoel P, Dequesne G, Feiron C, Lobet Y;
PI Poolman J, Thiry G, Thonard J, Voet P;
XX WPI; 2001-138654/14.
DR New isolated polynucleotide useful for outer membrane vesicle
PT preparation from Gram-negative bacterial strain for vaccination of
PT microbial infections -
PS Claim 46; Page 81; 128pp; English.

The invention relates to a genetically-engineered outer membrane vesicle (bleb) preparation from a Gram-negative bacterium for use as a vaccine. The blebs of the invention are improved with respect to their immunogenicity and toxicity by the introduction of one or more genetic changes to the chromosome of the bacterium from which the blebs are derived. The changes made include the upregulation of protective antigen expression, the downregulation of immunodominant non-protective antigen expression, and genetic changes which result in detoxification of the lipid A moiety of lipopolysaccharide (LPS). The invention also encompasses modified Gram-negative bacterial strains from which the bleb preparations are made, a vector suitable for performing recombination events (for the generation of the modified bacterial strains), bacterially-derived nucleic acid sequences used in such a vector, and an immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole cell vaccine suitable for paediatric use. The bleb preparation is useful in the manufacture of a medicament for immunising a human host against a disease caused by infection of one or more of the following: *Neisseria meningitidis*, *Neisseria gonorrhoeae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, *Pseudomonas aeruginosa*, *Chlamydia trachomatis*, and *Chlamydialata pneumoniae*. The invention may also be used to provide immunisation against the influenza virus. Bacterially derived nucleotide sequences of the invention are used in the performance of homologous recombination events up to 1000 bp upstream of a bacterial chromosomal gene in order to either increase or decrease expression of that gene. Immunoprotective and non-toxic Gram-negative bleb, ghost, or killed whole cell vaccines are more immunogenic, less toxic and safer, and are particularly useful for paediatric use. The present sequence represents a specifically claimed *Neisseria meningitidis* nucleic acid sequence.

Sequence 731BP; 226 A; 169 C; 194 G; 142 T; 0 other;

Alignment Scores:	
Pred. No.:	1,24
Score:	83.00
Percent Similarity:	35.06%
Best Local Similarity:	21.26%
Query Match:	6.26%
DB:	22
Length:	731
Matches:	37
Conservative:	24
Mismatches:	71
Indels:	6
Gaps:	42

US-10-008-355-2_COPY_522_712 (1-191) x AAF91391 (1-731)

Oy	2	LysSerValIleAlaIleAlaIleArgAlaIleGlnIleLaspIleMetAlaAsnAlaTyrAla	21
Dd	61	AAAAACAGCAAGATCCGGATCCAGCGCAGGAAGAAAGCATGTACAAGTATGCTAAAACG	120
Oy	22	IleGluTyrGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAla	41
Dd	121	GACACAGTTGGACAAAGATATTCTCTCCAAAGC-----GAGCGGCACCATTGAAAAAAG	174
Oy	42	LeuProSerAspAlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu	61
Dd	175	ATTCCAAACGACAAACAGTGCTTATCCG-----	204
Oy	62	ProGlnAspGlyAlaIlePTrTyrAsnTyrHisThrThrGlyLys-----GlyVal	77
Dd	205	-----GGGTCTTGGTAGGGCCATAATGCCAACAGCGCACAGCTGAGCGGCATGCT	255
Oy	78	LeuGluTyrGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeu	97
Dd	256	TCCGATTAAAGAGGGCGGCACAGGCGGACTTTACTGTG-----	294
Oy	98	PheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyLeuHisIleAlaPheLeu	117
Dd	295	-----AATTTCGGTAACAAAAAATATTAAGCGACAGTTA-----	327
Oy	118	SerAsnAspIleThrGlyLysSerGlySerProValPheAspLysAsnGlyArg	137
Dd	328	-----ACCGCTGACAAACAGCGCGGCGCAACCTTACCATGTGGCGCAT	372
Oy	138	LeuIleGlyLeuAlaPheAspGlyAsnTrpGluLametserylGlyAspIleGluPheGlu	157
Dd	373	ATTGAGGGCAACGGTTTTTCCGGTAGC-----GCGAAACCTGCTACCTCAGGTTTTGAT	426
Oy	158	ProAspLeuGlnArgThrIleSerValAspIleArgTyrVal	171
Dd	427	CTCGATTCAAGCAATTAACACCCGACGCGCTAAGSCATATATC	468

Search completed: December 21, 2002, 02:00:26
Job time : 749 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 17:31:45 ; Search time 2836 Seconds
(without alignments)
1960.027 Million cell updates/sec

Title: US-10-008-355-2_COPY_522_712
Perfect score: 1003
Sequence: 1 SKSVTAARAIAIQADAMANAY.....LFMIDKMGCPRLQELKLI 191

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=xlh
-Q/cgqn2.1/USPTO.spool/US10008355/runat_17122002_112704_16920/app.query.fasta_1.327
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR.SCORE=DCT -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355_6CGN.1.1.1616_@runat_17122002_112704_16920 -NCPU=6 -ICPU=3
-NO_XLPPX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hg_hum:*
31: em_hg_inu:*
32: em_hg_other:*
33: em_hg_mus:*
34: em_hg_pin:*
35: em_hg_rtd:*
36: em_hg_mam:*
37: em_hg_yrt:*
38: em_sy:*
39: em_hggo_hum:*
40: em_hggo_mus:*
41: em_hggo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	334	33.3	10689	1	AE004008	AE004008 Xylella f
C 2	176.5	17.6	253217	2	AC016590	AC016590 Homo sapl
C 3	175	17.4	10811	1	AE011732	AE011732 Xanthomon
C 4	101.5	10.1	204050	1	AL646073	AL646073 Ralstonia
C 5	93	9.3	11581	1	AE007581	AE007581 Clostridi
C 6	91.5	9.1	79718	8	AF012H2	AL807577 Aspergill
C 7	91.5	9.1	180688	9	CNS01RC9	AL157871 Human chr
C 8	91.5	9.1	187710	9	CNS01DVI	AL158388 Human chr
C 9	90.5	9.0	1554	14	AIU52600	U52600 Avian infec
C 10	90	9.0	2143	6	AA7465	AA7465 Sequence 38
C 11	90	9.0	2163	1	NMTBP2163	Z50731 N.meningiti
C 12	89.5	8.9	2794	9	AK023006	AK023006 Homo sapl
C 13	89.5	8.9	2908	10	BC021506	BC021506 Mus muscu
C 14	89.5	8.9	2956	9	BC027713	BC027713 Homo sapl
C 15	89.5	8.9	3048	9	BC009988	BC009988 Homo sapl
C 16	89.5	8.9	3058	10	BC027844	BC027844 Mus muscu
C 17	89.5	8.9	3158	9	BC002564	BC002564 Homo sapl
C 18	89.5	8.9	3513	9	HS47509	AJ007509 Homo sapl
C 19	89.5	8.9	3514	9	AK022863	AK022863 Homo sapl
C 20	89.5	8.9	10862	6	AR124177	AR124177 Sequence
C 21	89.5	8.9	10862	6	AX002387	AX002387 Sequence
C 22	89.5	8.9	10862	14	AF052437	AF052437 Yellow fe
C 23	89.5	8.9	10862	14	AF052438	AF052438 Yellow fe
C 24	89.5	8.9	10862	14	AF052439	AF052439 Yellow fe
C 25	89.5	8.9	10862	14	AF052444	AF052444 Yellow fe
C 26	89.5	8.9	10862	14	AF052445	AF052445 Yellow fe
C 27	89.5	8.9	10862	14	AF052446	AF052446 Yellow fe
C 28	89.5	8.9	10862	14	FLXFI7DG	X03700 Yellow feve
C 29	89.5	8.9	10862	14	YF017066	U17066 Yellow feve
C 30	89.5	8.9	10862	14	YF017067	U17067 Yellow feve
C 31	89.5	8.9	10862	14	YF021055	U21055 Yellow feve
C 32	89.5	8.9	10862	14	YF021056	U21056 Yellow feve
C 33	89.5	8.9	10862	14	YF021057	U21057 Yellow feve
C 34	89.5	8.9	10862	14	YF054798	U54798 Yellow feve
C 35	88	8.8	12602	3	DEU40653	X15062 Yellow feve
C 36	87	8.7	315079	1	MP0LM03	U40653 Drosophila
C 37	86.5	8.6	3832	1	CAJFLAAB	AL445565 Mycoplasma
C 38	86.5	8.6	7756	1	AF0202168	J05635 Campylobact
C 39	86.5	8.6	10760	14	AF094612	AF0202168 Campyloba
C 40	86	8.6	9293	1	AE011022	AF094612 Yellow fe
C 41	86	8.6	10590	1	AE006285	AE011022 Methanosa
C 42	85.5	8.5	1003	8	AY091427	AE006285 Lactococc
C 43	85.5	8.5	1156	8	AF114386	AY091427 Arabidops
C 44	85.5	8.5	1177	8	AY056227	AF114386 Arabidops
C 45	84.5	8.4	1731	1	AF050190	AY056227 Arabidops

RESULT 1

ALIGNMENTS

AE004008/c	10689 bp	DNA	linear	BCT 15-JUN-2001
LOCUS	Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.			
DEFINITION	Xylella fastidiosa 9a5c, section 154 of 229 of the complete genome.			
ACCESSION	AE004008 AE003849			
VERSION	AE004008.1 GI:9106961			
KEYWORDS	Xylella fastidiosa 9a5c.			
SOURCE	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;			
ORGANISM	Xylella fastidiosa 9a5c.			
REFERENCE	1 (bases 1 to 10689)			
AUTHORS	Simpson,A.J., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R., Bueno,M.R., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,V.C., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., Guerrier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P. and Marino,C.L.			
TITLE	The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing and Analysis			
JOURNAL	Nature 406 (6792), 151-157 (2000)			
MEDLINE	20365717			
PUBMED	10910347			
REFERENCE	2 (bases 1 to 10689)			
AUTHORS	Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M., Alvarenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S., Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S., Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M., Carrier,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R., Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E., Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,V.C., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C., Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H., Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L., Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Lairet,F., Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A., Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N., Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L., Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C., Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A., Nascimento,A.L.T.O., Netto,L.E.S., Nhani,Jr., A., Nobrega,F.G., Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C., Palmiter,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira,Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V., de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V., Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva Jr.,W.A., da Silveira,J.F., Silvestri,M.L.Z., Siqueira,W.J., de Souza,A.A., de Souza,A.P., Terenzi,M.F., Trifil,D., Tsai,S.M., Tsunaka,M.H., Vallada,H., Van Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M., Meidanis,J. and Setubal,J.C.			
TITLE	Direct Submission			
JOURNAL	Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil			
FEATURES	Location/Qualifiers			
source	1..10689			
gene	/organism="Xylella fastidiosa 9a5c"			
CDS	/db_xref="taxon:160492"			
	/clone="9a5c"			
	/complement(119..367)			
	/gene="XF1871"			
	/complement(119..367)			
	/gene="XF1871"			
	/note="hypothetical protein: identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"			
	/codon_start=1			
	/transl_table=11			
	/product="hypothetical protein"			
gene	/protein_id="AAF84677.1"			
CDS	/db_xref="GI:9106962"			
	/translation="MTMTLKEDKKEIIDEQRIQIAKLISERAKIQAEIHENATQKL MKETIKLTESKWPVYVSSSLIAGATATLTFLFH"			
	complement(430..618)			
	/gene="XF1872"			
	complement(430..618)			
	/gene="XF1872"			
	/note="hypothetical protein: identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"			
	/codon_start=1			
	/transl_table=11			
	/product="hypothetical protein"			
	/protein_id="AAF84678.1"			
	/db_xref="GI:9106963"			
	/translation="TMEARATAPQGGIASDPQRTIDAPKTKLRRKRYPTTSASI IALRVLTQROOKRASP"			
	941..1288			
	/gene="XF1873"			
	941..1288			
	/gene="XF1873"			
	/note="similar to GI1479340 (percent identity: 36 %/query alignment coverage: 87.8 %/subject alignment coverage: 90.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"			
	/codon_start=1			
	/transl_table=11			
	/product="conserved hypothetical protein"			
	/protein_id="AAF84679.1"			
	/db_xref="GI:9106964"			
	/translation="MSKNIPNQLITGMAGFQINGKLVTPERTTEEMQLMLSLT CLARECKMEERATAPQGGIASDPQRTIDALQTVSHKRHAANETASIIOLRO VLIRQOKRASP"			
	1351..1662			
	/gene="XF1874"			
	1351..1662			
	/gene="XF1874"			
	/note="hypothetical protein: identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"			
	/codon_start=1			
	/transl_table=11			
	/product="hypothetical protein"			
	/protein_id="AAF84680.1"			
	/db_xref="GI:9106965"			
	/translation="MTMTLKEDKKEIIDEQRIQIAKLISERAKIQAEIHENATQKL NAHNENAKRQENAKRQIKIEETQKSKESQWPVYVSSSLIAGATATLTFLFH"			
	complement(1839..2918)			
	/gene="XF1875"			
	complement(1839..2918)			
	/gene="XF1875"			
	/note="similar to SP1P03626 (percent identity: 36 %/query alignment coverage: 88.0 %/subject alignment coverage: 105.0 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"			
	/codon_start=1			
	/transl_table=11			
	/product="phage-related protein"			
	/protein_id="AAF84681.1"			
	/db_xref="GI:9106966"			
	/translation="MLNLVTGAPGNGKTLVAVDWLIRQIEIDKSLVSGAVPRSYTPD IEGFDVEAVRLTGVVOSAPEDRTPPGQSVIYDEAHMFPAQRGRSDPPVCDL DTHHGGDMLFVOWPPTKIHHLRLVGEVHLNBRAMGLQTAGLRMSRDDPYDI HQRKAREEVKPFKDRALATASLTHTVSKRPFKKMSALSCVTCSTGIVPFIH TYSFAHLSSEASVAGAGQASLRAAPSLSSSSLSVSGMRTAVLFTESAPVLSCV SSRSRCRCFTDQVIDMSVVECRLLASPLPFVYHAYVYSSSSAAVSSSPSSPS SSSVSLSSSVSAPILGSSSVPSH"			
	complement(2918..3202)			
	/gene="XF1876"			
	complement(2918..3202)			
	/gene="XF1876"			
	/note="similar to SP1P03625 (percent identity: 36 %/query alignment coverage: 85.1 %/subject alignment coverage: 85.1 %)			

Qy	8	AlaAgaGaaIaIegInaIaSpaIaMetuAaIaSnAaIaTrrAaIaIaIegIuLySgIyIaSrG	27
Db	8864	GGCGCTCGAATCTACTTCAAGCCCTTACTGTGAC--TCAACACAAAGATCATGGCAAA--	8811
Qy	28	LeuPhepheaIaGlyLeuAaRgGluMeteyrProGlyaTgaIaIaLeuProSerAaSpaIaSn	47
Db	8810	-----TTTCTTTTACCA-----GAGCCCAAC	8790

QY	48	PhetrimetAlaMeSerTyrGILSerLleYsGLYrYrGluProGlnAspGlyValantrp	67
Db	8789	TCATCATTTACGTAATTACTCTTGCGGCACGCTGAAGGCTATTACCCCAAGAATGGCGTCGAG	8730
QY	68	TyrAsnTyrHisThrThrGILyLysGLYValLeuGlnLysGlnAspProLysSerAspGlu	87
Db	8729	TACACACCGTTTACGACACTCAGAGGTGTCAATGGCGAAG--AATACAGGCGTGCAGCA	8673
QY	88	PhelaValAlGlnGluAsnIleLeuAspLeuPheArgThrGlyrYsAsnTyrGlyArgTyrAla	107
Db	8672	TTTCGATTACCGCAGAGTCCTATTAAATCCGATTAAGGCCAAAGAGCTACGCCAATTTACCC	8613
QY	108	GluAsn-----GILyGlnLeuHisLleAlaPheLeuSerAsnAsnAspLleThrGlyGly	125
Db	8612	GATCAGCGCTATTGGCACCTGCTCCCTGTGAATTTTCTATCTGTGACTTGACATTACTGGCGGC	8553
QY	126	AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGly	145
Db	8552	AACCTCGGATCGCCAGCTCTTGATGCCACGGAATAATGTGGGGTTTGACATTTGACCGGT	8493
QY	146	AsnTrpGluAlaMetSerGlyAspLleLeuPheGluProAspLeuGlnArgThrIleSer	165
Db	8492	AATTGGAGCTGTGCAGTAGTAATTGGGTGTGATCTGTGATGACTCGAACAATTTGCA	8433
QY	166	ValAspLleAlaMetSerGlyAspLleLeuPheMetLleAspLysTrpGlyGlnCysProArgLeu	185
Db	8432	GTTGATTACTCGTTATGTGCTGACTGATGATCTATGACTGTAAGTTGACACCGACACCATCTACTA	8373
QY	186	GlnGluLeuLysLeu	190
Db	8372	AAGCAATTGAATTGG	8358
RESULT 2			
LOCUS	AC016590/c		
DEFINITION	Homo sapiens chromosome 19 clone CTD-3220F14, WORKING DRAFT		
ACCESSION	AC016590		
VERSION	AC016590.6		
KEYWORDS	HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 253217)		
TITLE	DOE Joint Genome Institute.		
JOURNAL	Sequencing of Human Chromosome 19		
REFERENCE	2 (bases 1 to 253217)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	DOE Joint Genome Institute.		
JOURNAL	Submitted (04-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	On Apr 20, 2001 this sequence version replaced g1:7711568.		
	-----Genome Center		
	Center: Joint Genome Institute		
	Center Code: JGI		
	Web site: http://www.jgi.doe.gov		

	Project Information		
	Center Project Name: 940643, BC905667		
	Center clone name: CITB-EL_3220F14		

	Summary Statistics		
	Consensus quality: 207648 bases at least Q40		
	Consensus quality: 225569 bases at least Q30		
	Consensus quality: 232428 bases at least Q20		
	Estimated insert size: 250510; agarose-fp estimation		
	Estimated insert size: 249417; sum-of-contigs estimation		
	Quality coverage: 9.03 in Q20 bases; agarose-fp estimation		
	Quality coverage: 9.07 in Q20 bases; sum-of-contigs estimation.		
	* NOTE: This is a 'working draft' sequence. It currently		

REFERENCE AUTHORS	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; 1 Xanthomonas 1 (bases 1 to 10811) da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J., Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locati,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Mench,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kittajima,J.P.
TITLE	Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
JOURNAL MEDLINE PUBMED	Nature 417 (6887), 459-463 (2002) 22022145 12024217
REFERENCE AUTHORS	2 (bases 1 to 10811) da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardoso,J., Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locati,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J., Mench,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kittajima,J.P.
TITLE	Direct Submission
JOURNAL	Submitted (28-NOV-2001) Departamento de Bioquimica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil
FEATURES source	Location/Qualifiers 1. 10811 /organism="Xanthomonas axonopodis pv. citri str. 306" /strain="306" /db_xref="taxon:190486" /note="pathovar: citri" complement(293..1315) /gene="tdh" /note="XAC1022" complement(293..1315) /gene="tdh" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="threonine 3-dehydrogenase" /protein_id="AAM35905.1" /db_xref="gi:21107162" /translation="MKALVKKRANKGIMLEOVVPTPGPNEVLIKLEKTAICGDTLHI YLDMWSORTTEPGTLTIGHEFYGRAELISANTGYOIGORVASEGHIVCGHNRNCG RPHLCPTNVGICVNNNGAFEMVNPASNPIMPDIQISELAARFDPIGNANHCALF DVIGDVIITGAGPIGIIAGICKIHGARNVVYVNDVDFRLKLAADKCATRVVNSKT SLKDVADLHMEGFDVGLMEGNSRAFDMDLDCMYHGKIAMLGIMPRGAGCDMDKII FKGLTVGGIYGRKMYETWYKMTQLVLSGFPLHKVLTLPIDDFQKGDIMEGRKAGK VILSWN" complement(1840..4260) /gene="fecA" /note="XAC1023"
gene	
CDS	
gene	complement(1840..4260) /gene="fecA" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="tonB-dependent receptor" /protein_id="AAM35906.1" /db_xref="gi:21107163" /translation="MFHEPLTVCVAKRRLCYRPPSAVPEDSPTASRRHYVPLAL ALATLPIVTSQATPADPAVAAPADALADAAATDADANVDSQSTRVVISQ ODIEQLTPGSSAFKAEKLPQVQFQADPFGEYENSTGYTLHGFDQSRIGTLDLIPL GMSYGVNGHITPAITISENYSVEIKQAGALGTASTNTNGMGCEYSDAPOTTPG ARLVOTSDATRRFVRADRDGRLSAVLSANASTDKMKYGDOTSEANLKTIV OMGDNRLSLFVDRSRREKDYMDLSLTQSRALGNVYVLDPMATVAPOMARVONTG ATSGVANGYPSIALPLPDDISYMDLSYTAGGRLRDNLAGLSGTREGCATLDASGY HNRREGVWTFYVTSKQIPVSKRTTYGLDFREGTSALKMSWGHOLEGAMAEHA RTTQGRNFALGNEGYTSLVNYEATQEPFRDLQRYTQRMLYVQDTVRLDLRLT LNYGAKALSTTRRAQSLVPTTALAGRTIAEDNFPOLGINKYKIDERDLVASYKNI AAFGTTPATSOAFDRSRSTLEPQSOVQVGYVQAOQLSDADAFETFSNLLT TSPCTAVQTCASILSNCGAVHAGADLAIMRPTEGLSLMINSLSNRSRYDDYLNNG VYATSGKRVGICPALMFSSSASVQIGNRLRDDCYVYKRYITPELNDQVSYSTLENA GARYDFGRVGVADVALLALNISLTDKRIFASTGTNGVYASDPDSYNOTWVGAPROY FMSLDVKE" 4369..6450 /gene="XAC1024" 4369..6450 /gene="XAC1024" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="non-hemolytic phospholipase C" /protein_id="AAM35907.1" /db_xref="gi:21107164" /translation="WTSSSSRRDFLKRVAALTAAGALPSSIGRALALPANKRTGLRD LEHVILMOENRSEDFHFGALGVEQDPDALQRLDHAWVSQTPDGRRLVPAFD SONTCAPYKISLHSMKAGHODPRAMEYDAMVYKELMNGYGRDIPYVHLAD AFTCIDGYCSLHGPTNPNRMVLTFTGSGPSGVNNGAAYANADGNMTAAMARPCG VYALQMTVYAOQLQAGVDWRYQGYDNPFGCSLAFSEHYRLRLTRDDERYLRAKCP GSTAGNAATTOEHLVAAIADVRANRLPQVSWVLPVYACEHPAPPAVESTVARI IDALTANPEWAKTALIIYNDENDGFPHVAPLPALDARMGRSNVDHGEYDVAPI GLIGRPMVLVSPWTRGCVNSQVDFDTSVRLERRGVAFENISPMRRAVSGDLS VFDEFRPDDLSALSPVDVDRAPRAAARDPLPAPAAATPBPQEGCORPARLPYA LOVHARVOGAAVOLOFVNSGAAANVYNSAAGCGPWYTVLPGTOLDAPICATH EGAYALRYHGENGFREFRAGOVPPRASAPSAAPWAEARQEGDALVLEINACQACTVO LRADYADPSARTLSLAAGQRETRIALAASDAHWYDLVEQPSAFERRRAGHLETR PSRSDPAFGRY" 6617..6766 /gene="XAC1025" 6617..6766 /gene="XAC1025" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAM35908.1" /db_xref="gi:21107165" /translation="MRPNLFAVSVLATLAVAASAGQGEGMVVPQQLPEIAGPQOAGL RLSPR" 6791..7054 /gene="XAC1026" 6791..7054 /gene="XAC1026" /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="AAM35909.1" /db_xref="gi:21107166" /translation="MGAVVALGCGTASFVSPQGLVTVTNHHCAGYAIQULNSTAQRNLIK

gene 7236
/gene="XAC1027"
7051..7236
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA035910.1"
/db_xref="GI:21107167"
/translation="MMDAOGKLVGLAFDGNMESVSSNMVIFDPAMTRMIAVDGRYLRLMI MTEVAPAPOLKELEGLVR"
7329..7973
/gene="pgma"
/note="XAC1028"
7329..7973
/gene="pgma"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="phosphoglycerate mutase"
/protein_id="AA035911.1"
/db_xref="GI:21107168"
/translation="MRILLARHGEMPMNAEGRYOGIDIPISPGEGQARALGERTLS LQIDRAVASPLSRQATKAKALGASRQALLQTDADLOIANGEMGLASELNDKQDPA RLKRAKEEDVYLMFGESLROVLDRSRGLMRADGLGADHTLLVADHVNRYTLC KILGLPLRLMSFRQAPTTLNLLEGGDVEHLEVLNDCAHHTPEFGAKHRAL"
8242..9540
/gene="fo1c"
/note="XAC1029"
8242..9540
/gene="fo1c"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark; dihydrofolate synthase"
/codon_start=1
/transl_table=11
/product="folypolyglutamate synthase"
/protein_id="AA035912.1"
/db_xref="GI:21107169"
/translation="MNATDSLSAWLAYIEQHPSAIANGLEREVAARLQIAAPAKH VIVVGTMGKSTVAFIEAIGRAAGMKVGAATSPHLRLYNEVRIDGQASDAPQVDA FAAYEARGDALTAFERFGTLAALMLFQGSALVELALEIGGRIDANVINDSDVAVI

Alignment Scores:
Pred. No.: 1.27e-08 Length: 10811
Score: 175.00 Matches: 32
Percent Similarity: 75.00% Conservative: 16
Best Local Similarity: 50.00% Mismatches: 16
Query Match: 17.45% Indels: 0
DB: 1 Gaps: 0

US-10-008-355-2_COPY_522_712 (1-191) x AE011732 (1-10811)

QY 125 GYAASerSergISerProvalPheasplysAsnGlyArgLeuIleGlyLeuAlaPheasp 144
|||||
DB 7033 GGCAACTCCGGTTCGGGTGATGACGGCGCAAGGCAAGCTGTCGGCTGCGCTTCGAC 7092
QY 145 GYAASrTrpGluAlaMetserGlyAspIleGluPheGluProaspLeuGlnArgThrile 164
|||||
DB 7093 GGAATTTGGGAGTCGGTGAACGACCAACTGATCTTGACCCGGCAATGACCCGCAATGATC 7152
QY 165 SerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGlnCysProArgLeu 184
:::|||||
DB 7153 GCGGTGACGCGCGCTTATCTCGGTGATCATGACGACGAGTGCCTCGGCCACACGCTG 7212
QY 185 IleGlnGluLeu 188
:::|||||
DB 7213 TTGAAGAACTC 7224

RESULT 4
AL646073/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Al646073 204050 bp DNA linear BCT 07-DEC-2001
Ralstonia solanacearum GM11000 chromosome, complete sequence;
segment 17/19.
AL646073 Al646052
AL646073.1 GI:17429991
Ralstonia solanacearum.
Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 204050)
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choisme,N., Claudel-Renaud,C., Cunac,S., Demange,N.,
Gaspin,C., Lavié,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Sigulier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 204050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France. Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex. Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers
1..204050
/organism="Ralstonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
complement(115..1278)
/gene="dgt"
/note="RSC2968: RS01329"
complement(115..1278)
/gene="dgt"
/EC-number="3.1.5.1"
/function="small molecule metabolism; central intermediary
metabolism; nucleotide hydrolysis"
/note="Product confidence : probable
Gene name confidence : probable
Predicted by Codon-usage
Predicted by Homology
Predicted by Framed
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE DEOXYGUANOSINE TRIPHOSPHATE
TRIPHOSPHOHYDROLASE PROTEIN"
/protein_id="CAD16677.1"
/db_xref="GI:17429992"
/translation="MNDRLLDLPAPFDLEGLAPYAAHAOSRGRVNAEPSTSRTER
ORDRRITHCARFRRLKRYQVFNHGGDLFRRTYSLVAQIARSTARSRLNEDL
VEAISLADLHCHTFPGAGDALNDCKAPYCGFPHNLSLIVRLERRGGFGNLN
TFETREGILKHCSSNNNAATLGDLRRELEGGQPSLEQOLANLADENVVNNHIDIDGLR
SGLTTLDELEGVPMARHREAEAFPGVGRRMINTIRIINALIVDLTACTRAAI
LGAPOSIDIVRAAPLVAFSEPMDEARVRLFLPNLYRHYLVMMMAAKARRTIED
LFAVLEDPRLPLPOYARDPADQPRMIAHYIAQMTBRYAIKHKRIFAVEVL"
complement(1298..2404)
/gene="arob"
/note="RSC2969: RS01328"
complement(1298..2404)
/gene="arob"


```
/EC_number="4.6.1.3"
/function="small molecule metabolism; amino acid
biosynthesis; chorismate"
/Note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE 3-DEHYDROQUINATE SYNTHASE PROTEIN"
/protein_id="CAD1678.1"
/db_xref="GI:17429993"
/translation="MTIVVDLGERAAPHIGTGILSOAELFAPHIRGTRAVIYNET
VAPLYARVEAARSLIGKTVDMVYLPDGESEFTWETLIRFDALLASGADRKTIVAT
GGYIGDMGTGFAASVYMGVGFIOVPTLLSQVSSVSGKGINHPCLKNMGAFHOP
OAVLADITLRLTPPRELAGMAEYIKGALADADYFAMIRHIALGLNACDAMAGAP
VRSVQIKAAVVAODERESGLRALINEGTFGHAIEAGLGWELHGEAVCGAMAA
DLSHRLGFDIDITRNRYALTALRAANLPVAPADLGVARFIDIMRDVKKAGEIKRVLL
RKLGAFTVTPDTRATLQHAVALRPTAPVA"
/complement(2401..2973)
/gene="arok"
/Note="RSC2970; RS01327"
/complement(2401..2973)
/gene="arok"
/EC_number="2.7.1.71"
/function="small molecule metabolism; amino acid
biosynthesis; chorismate"
/Note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE SHIKIMATE KINASE I PROTEIN"
/protein_id="CAD16679.1"
/db_xref="GI:17429994"
/translation="MSVSNVFFVGLMGAGKTTGGAARVRLDPPFSDHEARAGV
RVPVIFHEGEMGFREDTQMIDELTAHGVAVATGGAVLRPENRAFLREGVITYL
RANPDILRLTRHDKNRPLQETENRRLREELHAIRDLRYEVAHFVETGKPIVAOL
VNMVLMQLEVAGIVVPAASSPTSQSVRSQS"
/complement(3114..5258)
/gene="p110"
/Note="RSC2971; RS01326"
/complement(3114..5258)
/gene="p110"
/function="structural elements; cell exterior; surface
structures"
/Note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE FIMBRIAL TYPE-4 ASSEMBLY SIGNAL PEPTIDE
PROTEIN"
/protein_id="CAD16680.1"
/db_xref="GI:17429995"
/translation="MTMRVKGGLVQAVRVARGVAVVNLISCLMWGQALAQQAITA
PASNAVEREASSTGEESTVITVLKDTPAQKVERSTQOPATIALIDFGTAPASQRAN
YQYGGKLLRAASVQIDGTRVYLDLAQSQYSEIKRNOYVLTIGAAPATASATPPT
FAAPAPATGVSERANIDFRGELIAGRVVVDLSTNSAINIAQGGONLVDFEGAT
LPQSLRRFVDVSEFSTPVQAMRATNGGAVLIPRGNMOWSSYOTDQFVFEVAPT
KEDPKLISGPGYGERMSLNQNDISLLOVDFPNINLVTSDSVTGTSILAKD
VPMQALQIIVDSKGLASRGNVIVVAPRNEALKEKAELESQOQVYELDELRSQV
RLNTRQADVRNMLGTGTLGGGAASRLISKRSGLISDARTNQLFVSDIPSKLEEV
QAFLLKIDIPVQVMIETARIVADTFESRNGAKIGFAKTNGAGYGTIVSNVSPVT
KDATWNSAAVSLPANGINGVNAASVAVSLFNAGARFLAELSLALADGRKTIISPP
```

```
RVATADNIKALIEQCTEVPYQQAATSSGATSVSFKKANIKLEVTETKIPDPGNVFLADV
NDSLIGQTNGFAINTKQVOTQVLYVGGTVVIGITVQNERFDVKNVPLGLDIPVL
GNLFSTAKTNDRTRELVETLPRVYLDQLSLK"
/complement(5255..5800)
/gene="p11p"
/Note="RSC2972; RS01325"
/complement(5255..5800)
/gene="p11p"
/function="structural elements; cell exterior; surface
structures"
/Note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE FIMBRIAL TYPE-4 ASSEMBLY LIPOPROTEIN"
/protein_id="CAD16681.1"
/db_xref="GI:17429996"
/translation="MKGVMTROYTRLIHSPLCLAVVLAACGASADELEOMNOQA
RSTVSTRAPLPEPKVAPREYTTKOTDPDSYQVVAELSRQNSPEPGRHREPLED
VPLEAFKILGTMRKGESEAVLSQDKTQHVHVGQYIQQNGRIVIGIDQELTLELV
REGTAEMKEMTTLKVGQAG"
/complement(5797..6462)
/gene="p110"
/Note="RSC2973; RS01324"
/complement(5797..6462)
/gene="p110"
/function="structural elements; cell exterior; surface
structures"
/Note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence-not_experimental
/transl_table=11
/product="PROBABLE FIMBRIAL TYPE-4 ASSEMBLY MEMBRANE
TRANSMEMBRANE PROTEIN"
/protein_id="CAD16682.1"
/db_xref="GI:17429997"
/translation="MTMALSTELSLLELVGQFGLNLEPPTMOLAPILPMIILAVL
VYVIGMOPYSGKILDERDAKRTQEQOLSAVOSKIAQVYVNEALRKQIEVERAKL
ELQLPNREMDALLADVNHAGIARSLITDLRPSQAVYKPYTAELRPSVYKXNRTHDM
ALFEADVAALSKRIVTLRNIALTGKDSMWEARAVRALDPPDQAAQRAAAVYKG
DAK"
/complement(6459..7091)
/gene="p11N"
/Note="RSC2974; RS01323"
/complement(6459..7091)
/gene="p11N"
/function="structural elements; cell exterior; surface
structures"
/Note="Product confidence : probable

Alignment Scores:
Pred. No.: 51 Length: 204050
Score: 101.50 Matches: 33
Percent Similarity: 41.61% Conservative: 24
Best Local Similarity: 24.09% Mismatches: 51
Query Match: 10.12% Indels: 29
DB: 1 Gaps: 3

US-10-008-355-2_COPY_522_712 (1-191) x AL646073 (1-204050)
OY 16 MetAlaasnaLstyzalalecgluLysgLyAsrgLeuPhePhehlaeglyLeuArGlu 35
::: ||| ::| |||||:::
Db 190879 ATCACCACACTGCCACTGCTGCGCGCGGCAAGCAGCTGCTGCTACG----- 190832
OY 36 MetTyrProGlylArGalaLeuProSerAlaasnaPheThrMetArGmetSerTyrGly 55
:::||||||| ||::: |||
```

Db 190831 -----CGCGGCAATGCCAATTTCGGCGCGCTTCGAGTATCCG 190793
Qy 56 Ser1eLys-----GLYrGrLupProGlnasp 64
Db 190792 GACGTGACCGGACATGTGCGCAGCTGCCGATTCACATATCTCCCGTGACC 190733
Qy 65 G1yAlArPrTyAsnTyRHisThrThrGlyLysG1yValLeuGluLysGlnaspProLys 84
Db 190732 CTTGGCCCGGACGCGCTGTGACGGCCAGAGAGTGATGCGCATCCGATCCGCTC 190673
Qy 85 SerAspGluPheAlaValAlnGluAsnIleLeuAspLeuPheArgThrTyLysAsnTyRgly 104
Db 190672 GGCTCGGACCTGACCATCGACGAGGCTGATCTCGCTCGCGCACCGCATGACGCG 190613
Qy 105 ArgTyAlaGluAsnGlyInLeuHisIleAlaPheLeuSerAsnAsnaspIleThrGly 124
Db 190612 CCGCTGACG-----ACGCTGACACCTCGCGCGCCCATCTCGCGG 190574
Qy 125 G1yAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeu 141
Db 190573 GGCTCAGCGCGCGGCTGTTCATGACCAACGCGCGCTGATGCGCATC 190523
RESULT 5
AE007581
LOCUS AE007581 11581 bp DNA linear BCT 27-JUL-2001
DEFINITION Clostridium acetobutylicum ATCC824 section 69 of 356 of the
ACCESSION AE007581 AE001437
VERSION AE007581.1 GI:15023527
KEYWORDS
SOURCE Clostridium acetobutylicum.
ORGANISM Clostridium acetobutylicum
Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
Clostridium.
1 (bases 1 to 11581)
Molling,J., Breton,G., Omeichenko,M.V., Markarova,K.S., Zeng,Q.,
Gibson,R., Lee,H.M., Dubois,J., Qiu,D., Hiltl,J., Wolf,Y.I.,
Tatusov,R.L., Sabathe,F., Doucette-Stamm,L., Soucaille,P.,
Daly,M.J., Bennett,G.N., Koonin,E.V. and Smith,D.R.
Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum
J. Bacteriol. 183 (16), 4823-4838 (2001)
21359325
JOURNAL MEDLINE
PUBMED 11466286
REFERENCE 2 (bases 1 to 11581)
CHILDRESS,D., ZENG,Q. and SMITH,D.R.
AUTHORS Direct Submission
TITLE Submitted (24-JUL-2001) GTC Sequencing Center Production,
JOURNAL Flushing, and Bioinformatics teams, Genome Therapeutics Corp., 100
Beaver Street, Waltham, MA 02453-8443, USA
FEATURES
source
1. 11581
/organism="Clostridium acetobutylicum"
/strain="ATCC 824"
/db_xref="ATCC:824"
/db_xref="taxon:1488"
154. 819
/gene="CAC0653"
154. 819
/gene="CAC0653"
CDS
154. 819
/gene="CAC0653"
/codon_start=1
/transl_table=11
/product="Response regulator (Chey-like receiver domain
and HTH DNA binding domain)"
/protein_id="AAK78630.1"
/db_xref="GI:15023528"

CDS
813. 2354
/gene="CAC0654"
/codon_start=1
/transl_table=11
/product="Sensory transduction histidine kinase"
/protein_id="AAK78631.1"
/db_xref="GI:15023529"
/translation="MLKIKSRILALYSMLTALIIIFIISFCILINIIYINKRPILDSI
TKVDGRKESEELNSRNHILTKNFIYEGAEKGKIKKSMIRVDSNRKQYIIIGFYVEKKG
GVIDYVTEKREKRYONKTYONTLETFIRNKGKIIITILEDPDYSEDNLSKCYCTKNLKY
SKGQIFIDOLKNNNSGKLSIVTDKNSVIRIOKIDSVYDKMGELSNNNGEYKMTSG
EEFAVMHVLVRFYIIIFIILMITITFALSTYVFAKPLIEFTHKIKESNAE
EIKFEVYGKVNDEIYDLICAVNALMEKIKNSYDLOILNSYSAHELNSHILAKILO
LGDKINIEAVYIDKTTNTNDIMASPTKLPFNDENVDAIICAVYVDEYIEVFNIEL
KLPGNVETLIGKIGKELMERCVANILDNAIKFVDKDNKAYNEIYVLEDEKRSIIIRVYD
NGIGIDESKIDKIFEPYIGNSMRISTGIGLAVYKHIMDLHGRVLVSKKPEYSEISL
IFSR"
3450. 3653
/gene="CAC0655"
3450. 3653
/gene="CAC0655"
/codon_start=1
/transl_table=11
/product="Hypothetical protein, CF-26 family"
/protein_id="AAK78632.1"
/db_xref="GI:15023530"
/translation="MKLVNPGRLTLETFAYSGCHVCSSASVYKQASRGFDEVYDSC
ECDMVEYTNIIYESTYIGIKNN"
complement(3884. 4453)
/gene="CAC0656"
complement(3884. 4453)
/gene="CAC0656"
/codon_start=1
/transl_table=11
/product="Predicted transposase (3' fragment)"
/protein_id="AAK78633.1"
/db_xref="GI:15023531"
/translation="MFLSSNQIPIYIKPQYEEKMKRSFKNDISRRENMEYDAKTQY
TCHNNRKLSPSCIIRKSSAGSYTSEVYECIEDNCTLKPKKAKKRNKQMSKAP
IKKREISNNIKITELGTRLRNRSIOVAGAGVLSDEYEFRLITRGNKSNKTEPILL
CFGININKLSKTIQKRTOKYIHLKLS"
complement(4489. 5493)
/gene="CAC0657"
complement(4489. 5493)
/gene="CAC0657"
/codon_start=1
/transl_table=11
/product="Predicted transposase (5' fragment)"
/protein_id="AAK78634.1"
/db_xref="GI:15023532"
/translation="MLVTYKLTHTKNVNOFNNDIOLILPLNLENLIPEDDSVRLSYLLE
GLNFKLYKKAASSIGKSSAVERKIMFKITISAYSONIYSSKIKACKRODRNFWLLO
SFYKPDHATISRRKKTLSNEVIEDLTFQOVNLYAKKELELFENVFIDGRTENANR
YTFYWKVITYIKNEGKMFDKIIVALEVDLENLAKGFTIGKRELLIDIDINKILLOMLLEKE
KRNIETFGIGKRRKTIQKWTQELQSVYKERQEKYNLSKIFSKNSYSKTDNDATPMH
MKDHRNGQLKPAVYNQIIVDSEYVGVGVDNRNDIATLIPMITNMOKKLYIHTPM
"
6172. 7728
/gene="CAC0658"
6172. 7728
/gene="CAC0658"
/codon_start=1
/transl_table=11
/product="Pe-S oxidoreductase"
/protein_id="AAK78635.1"
/db_xref="GI:15023533"
/translation="MSISEQCVIEIDRYIDYFKKNTSQPNILAKLFKTDSSVYYVDTG
TNKVLTCGEYYSILGKIDNKPFEIRINDADKSGEYVLEWROAIEKEDILLRG
KEKFEKAGHEELDKLNYELDQITLETFKGLRCRYCIYNSVYHNRKNGKMDP
IEIAEKADIVANLHSGKREKGVYTFYGGEPVYNKRLKHSIEYSRKYIKDKKLLFSLT
SNLTLPTLATYLASIENLITLCSIDGPEIHDSYKRDINGNGTFRAIRAGKLLYLE
AFGDKANKAVMMNVFTPPYSIEKVNIEFTFINSLDWLPGLRISMVYPTSGVDDTP
PGCEKNYGMELIOMSESKYSELKSDGKIDMFNEFEVPHILRIHGRIRISAKSESGY

PVNGCIIPIGSRKLYVTVDGNGFOICERINGSPIGNVFNIGIDENKIKNEVIDEFINKSI
KDCSKRMAARLCNWCYSHCYTDGKIDMKREKRYCIRSRELRALSLVFYHKCAKEINPEK
LKYNKIEVS"
7765..19084
/gene="CAC0659"
7765..9084
/gene="CAC0659"
/codon_start=1
/transl_table=11
/product="Predicted Zn-dependent peptidase"
/protein_id="AAK78636.1"
/db_xref="GI:15023534"
/translation="MNKEDGKNKIKNTLKNKFAVYIYCNKKSKGKVALIYKAGS
LMEQGLPNGTAHFEHLICIDSNVAYLKAVGAYYEDNINLQGTNEQYVYEDCTIDN
IDEGSVYKNILSDCFIDRLMEKVKDLYEIDYRMKCANFGIONNLLPYIIGKNI
KDKEPLKIKICEKMEFOSKSPFDKMKWKPONSALICGDIANSIDEMKISYSESEK
NDNEKTIKYSTNCKPYGKRYVNNVNNKNEVKAQYIYIKKDTISLAKSDTILYLFESI
IKNTILDLTDKSGEIVFYSKVLPHNLKLVNLEKRSOSEVERLEFIIDSINKIQAK
GINENDFVKRYRIEFKDKLSYHKEYYQSEIIEKCIINNFLEYDELMTLSNEYEACN
LIVDNIKFEFEKNKIIISQELNNKLLIVNGQEFNLLEYMKNFSRSL"
9167..19370
/gene="CAC0660"
9167..9370
/gene="CAC0660"
/codon_start=1
/transl_table=11
/product="Hypothetical protein, Cf-26 family"
/protein_id="AAK78637.1"
/db_xref="GI:15023535"
/translation="MKLVNPGVGRDLTELETAYSCHCICISGSGSYTAKMKGHYGADAC
ECDDVEHATNSDYADTFGIRKNN"
9575..11458
/gene="CAC0661"
9575..11458
/gene="CAC0661"
/codon_start=1
/transl_table=11
/product="ABC transporter ATP-binding protein (multidrug
resistance protein)"
/protein_id="AAK78638.1"
/db_xref="GI:15023536"
/translation="MKYSNPKKLGSLSSSKNIIRIYVLLKMKSSLSFIATLITVL
NGIIVPIMNVISKYLDISVNAALKNSPIDSLNKLFVLEFVGIASVYIGRNTY
FSTIOSKLNNSISSELIKRLNEDLSYFENDEPNFLKENSQALKNVSTIOVNLTO
IVKNEFTLVGAIVIVFOLSPITIIICITISIMPLINIKIKYKOKNITSKREKRLA
HYLQWAMDYIPVKEIKRNRIGDMLEKTIISITSKRNLDOKKIEKKOLLINSIVTEN
TTSIGKITYIYFTLITLORLSIGSNMTISALTINEDSVKNTLDFALLYSNLTEN
LFEVLDLKLPLLNNSGKKEFINELRDSIEFRNVSFKYPSNKYALKNVSEKIEANQ
CAIVGLNCGKTTLLKILRLYDPTEGEIYDINIKENIESLYKGIYVODEMKY
PEDVKKNIGFENIDNLFELIKTAARKANAVTFIOGSKNEEDTKLOKMSNGVDLSL
GOMOKLAISRAPMSNACLLIIDEPTASIDAKSEYELFPDFKEMGEETSIISHSRST
VKMDKTIIVLDGELVEBOGNDLSIKSGLYAELYNMQADATLVNEDKNDIVSKDKLK
DIV"

BASE COUNT 4472 a 1147 c 1925 g 4037 t
ORIGIN
Alignment Scores:
Pred. No.: 13.5 Length: 11581
Score: 93.00 Matches: 37
Percent Similarity: 44.00% Conservative: 29
Best Local Similarity: 24.67% Mismatches: 56
Query Match: 9.27% Indels: 28
DB: 1 Gaps: 7
US-10-008-355-2_COPY_522_712 (1-191) x AE007581 (1-11581)

QY 48 ptherhMetaIrgMeSeTyrIglYserILeYsgLYTyTgclUProGlnAspgIaLaTTP 67
Db 1098 TTTGTATATGAGAAAAGCTGAGTAATTGATGTATATAAAGCAAGAAAGATAT 1157
QY 68 TTTAsTtYthiThr-----ThiGlyLysgLYaLleu 78
Db 1158 CAAAATAATATACCTCAAAATATTATTAGAACTATATTATTAATGCTAAAGTTATATA 1217

QY 79 GltYsGlnAspProLysSerAspgIupheAlaVaIGlnGlnAsnILeUasPleuPhe 98
Db 1218 TCACATATACCTTCCA-----GATGATTTATCA---GAGATATATCTGAGTAAGTCTAC 1268
QY 99 ArgTthrLysAsnTyrIGlyArgTyrAlaGlnAsnGlyGlnLeuHisILeAlaPheUeSer 118
Db 1269 GGTATTAATAAATTGA---AAGTATTTATCTAAAGCCAGATTCTTCATTAAAGATTTACTT 1325
QY 119 AsnAspILeThrGlyGlnLysAsnSerGlySerProValPheAspLysAsnGlyAlaGlu 138
Db 1326 AAAAAATAATGAGTGT-----AAACATATCTATGATACACATTAAGATGATACGTA 1379
QY 139 ILegLYLeu-----AlaPheAspgLYAsnTPrGlnAlaMetSerGlyAspIle----- 154
Db 1380 ATTAGAATCCAGAAATTTGATGATAGCTTTATGATGATAAATGTTTGTGTAACCTTTCAAT 1439
QY 155 -----GluPheGluProAspLeuGlnArgThrIleSer 165
Db 1440 AATAAGGATTTAATAATGACATCAGAGAAATTTGAAGCTGCGGTATGATGATGACTCTAT 1499
QY 166 ValAspILeArgTyrValLeuPheMetIle 175
Db 1500 AGGAGATTTATTTATATATATTTATTAAT 1529
RESULT 6
LOCUS AFA12H2/c 79718 bp DNA linear PLN 27-JUN-2002
DEFINITION Aspergillus fumigatus BAC AFA12F2.
ACCESSION AL807577
VERSION AL807577.1 GI:21627806
KEYWORDS 3-hydroxyisobutyrate dehydrogenase; FAD synthetase;
isopentenyl-diphosphate delta isomerase; synaptobrevin
ATP-dependent RNA helicase.
Aspergillus fumigatus.
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
1 (bases 1 to 79718)
Harris,D.E., O'Neill,S., Knowles,D.G., Hall,N., Quail,M.,
Woodward,J.R., Denning,D.W., Anderson,M.J. and Barrell,B.
Direct Submission
Submitted (24-JUN-2002) The Sanger Centre, Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, UK
FEATURES
source
1..79718
/organism="Aspergillus fumigatus"
/strain="Af293"
/db_xref="taxon:5085"
join(2..1327,1382..4864)
/gene="AFA12H2.01"
join(2..1327,1382..4864)
/note="Similar to Saccharomyces cerevisiae hypothetical
269.9 kDa protein in Fkh1-met18 intergenic region Y11129C
SW:YIM9_YEAST (P40468) (2376 aa) fasta scores: E):
21e-67, 32.428 id in 1496 aa
Signal peptide predicted for Y11129C by SignalP 2.0 HMM
(signal peptide probability 0.606, signal anchor
probability 0.000) with cleavage site probability 0.431
between residues 18 and 19"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD37139.1"
/db_xref="GI:21627807"
/translation="IESHGIFPLCSQSRVRFAITVRLITPEPRALGKENTRIIRV
LEADHEILNVNDELQSVASERIRIOGRKRSASHNTLLELCSSEVSYSTLMSKYEPN
IIRISEFCPPAVTGTGREIVCARIOMKRTTIALDSPHPRYAPIDPAORALISRN
MTAETILQWKILYIMACTTINSVCAOSQSOALANNOHARKSKSGSQSQDRISSKSL
FAFVPLSAERSSTIRNAIVMLGSIINLNLFTPLSLOYATVTCNEAKTRIGAHHR
SPSSPRNNRKTDRLEFVTHYKLSHLREBEVYNDWIVNVLVYTKDLRIFLSDA
EVQNDWEFORLRFHYCGIMEELFEGINIKTKPSHWIIPESRVSASFSLMEDMGYSPNO
AQIAORENMRKFAMAHOSESGEFRTAAAMEIEKKNIRAAALSAMASICAGPISIT

ESGSVLQFDVGRMLSWIDIIENFLTSDKMHAIGRALKMLIHNKSHSLLEBRLAEMCH
VSEBKALESFEYEVQVLIENITDYPLGFENRILGAVTLTLCNOKEIREIMKSAKLIRIL
EEROKSSESLQDFDITSDTKTAAVYLKAOEFTSRRLAKOSHDLAPLTLESESAKLIRIL
GPDORNMVAAILPWOVTMELQIDPNGGTAKYMLANLFEITTRCGNITLIPNEQAL
MOLATLPHGNGVQVLDVFIITSLICERQONFEVAKOVYVFLSTPGSKVITFEFLM
QVPEKNMVOEKRDITPAPPDVKSLEYVADLGTVLPEVKNQAGLSIGVALLFEVLDM
APVTLLEAVVKLHVLLILMDHYMLTVQEQAREMLVHLIHELIAKIDDDAPATRO
GIEPVESESIRSDPKNVWEYEDNNKDEGDGSRVPLSMSTRQVNVFSEFAGSV
DLMAKEALNMATSCIPVRLACRSFOVFRCISTSLNPRMLADMLASNTIDEEADYO
TFSEMLITLTKIITISLAPADRLARPOLFMITCACTANTHETEPESIGMLEKPLDSY
DMSDPAYVTELIQOGPPKKEGGFDGLQNLVYKGLKSSSLNRTDVLIRISGLPNAL
IGNSDRLFTTLANNAHF.LHQFDPDVDPKTLARVTLTARAEGGCGCRPLASLGLA
NGQVKTENDFNHIIAEMRSYFFPQDVQSLIFLMLGTLNTTKMERVKIMKILCVLI
EIDMRPEVTHSHGPDLSPLRLLOTETELCPALAEVMDHIMVSGNPMERHHIRMSMAS
SSSSRAIKREYEKIQLSGIPEPTGMSIPMPATOSVVRHNHVAFYCAVEDPMEVO
ETVPEVEFHADEVNDSEFPKMRADTKMSMTQNMOMDIQKLSIDLSIDPEEPERIN
SGLDVMPASLRFTGATADISAHLYDOQTAPILRKS.LARTASTSFINGLAESPPN
QRFDSVGVTSPTLMPQVSGHLLRPHTHARSVTSFANNLYPPSSSNVQFSPLGSE
SAYLSDEDEFEDGLYEFEDRLTNKRPVYFHNSSQGSRYVDGSSLESMTSRRLTGG
AAANREKERQD.LIQAORAAVQTPANSRPVPEFLTGPTSNPTSPGQ"
/gene="Afa12H2.02"
/join(5738..5906,6086..6276,6472..7209)
/gene="Afa12H2.02"
/join(5738..5906,6086..6276,6472..7209)
/note="Similar to Neurospora crassa putative
3-hydroxyisobutyrate dehydrogenase g6g8.5 TR:Q9C1L3
(EMBL:AF309689) (338 aa) fasta scores: E(): 1.3e-65,
56.92% id in 332 aa
prosite profile:PS00895;3-hydroxyisobutyrate dehydrogenase
signature Confirmed by InterPro eMotif pattern
match :8e-5;codon 50-63"
/codon_start=1
/product="putative 3-hydroxyisobutyrate dehydrogenase"
/protein_id="CAD37140.1"
/db_xref="GI:21627808"
/translation="MSLTPTVSLRIRGTSVRKSLSSLSVSYPQLOARSARFTTLLQ
DATWGFILGQMGYLMANKLRAKIPATDTLLVRDVENATTRFEVETEEAKSSQAD
GTMAKEVQASAREVAEOSTVMTSLPPEPHKGVDFHSLIKGDLPALEERLFIDTST
IDTPASSEIANAISTRTGHFVDAMPASGCVGARGATSLTFMFGASGSGOLIERAV
LMLGKRAMHGLPGPAGVSGKLANNYLAIANNIAEAMNLGTRMGCLDPSIADIMNT
STGRCPMEVNNPVGVEYETAPANSRGYEGCGVGLMKKDLRLAIIAAKESGTPALAE
TAREYNSVEDEYRKQDSVYKMLQDSEQ"
/gene="Afa12H2.03"
/join(9141..9470,9608..9697,9745..9839,9901..10219,
10253..10567)
/note="Similar to Schizosaccharomyces pombe hypothetical
62.9 kDa protein c29a10.02 in chromosome II spbc29a10.02
or spbc365.18 TR:094381 (EMBL:AL034463) (567 aa) fasta
scores: E(): 3.1e-05, 33.85% id in 319 aa
pfam:PF00076;RNA recognition motif (a.k.a. RRM, RBD, or
RNP domain);0.16;codon 51-119
prosite pattern:PS50102;Eukaryotic RNA Recognition Motif
(RRM) profile:.110.326;codon 126-203"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="CAD37141.1"
/db_xref="GI:21627809"
/translation="MKTNGYMLQDLESITQOEPPIPAVPMAMTNPSEVTLAKLEN
QLISLTPNIRGFLPQTTDEMHSYAAARGKIDRCIAIVDLDTGLCPGVYKASSASAC
OLISLPDSKSNRSRLKLEDKTSTNICTNVTIWEADLRHHPPEHYHVESKISROE
KTGYSKEGFARPDTRETIAEKVYGFHNISNGVYLLIRFADTAOKKLLKOOSNERRA
YRAGYKNSYEVQGSTPSLSLQRLQOQTANHLSPPSVOYSYSPVCGSGWTMPATISIP
STRNDISRYOLMKNPASNMPPSSWSTNSPATWETTSVYHGRLYGRAIITPNNASASS
SKTYWPTPCVAKARGSGSPQKENNDVCLSPISCKEALMTSPSAL"
/gene="Afa12H2.04C"
/complement(join(11060..11087,11156..11243,11312..12053,
12210..12265))
/note="Afa12H2.04C"
/complement(join(11060..11087,11156..11243,11312..12053,
12210..12265))
/gene="Afa12H2.04C"

/note="Similar to Schizosaccharomyces pombe probable fad
synthetase spcc1235.04C SW:PADI_SCHPO (074841) (265 aa)
fasta scores: E(): 7e-31, 40.61% id in 261 aa
pfam:PF01507;Phosphodensinose phosphosulfate reductase
family;1.8e-11;codon 62-249"
/codon_start=1
/product="probable fad synthetase"
/protein_id="CAD37142.1"
/db_xref="GI:21627810"
/translation="MRESASPSRLSTNSDCAVQSFQIESHPDLSILAAVOOTRI
SLNVIATRLSKISELSITNGCKDCILVILITLASHHPHPEEGGLATIPATIAL
PLPDSPEAVEEVQSSRAYHIAIVKTEPPRTLLKSCFEHYLSLNSPKAIPIGTTR
TDPHGANLTHFDPDPSGMDPMLRHVPIVDMVAELTARIRLGLKYSGLYRGVTSIG
GTSITPMPKLRVEHSADPCGVGDQTSQVHYRAYELTEQOEERLGRILNNPSISQ
SSSLTFVHSAPQPDVGRCSNTELD"
/gene="Afa12H2.05C"
/complement(join(12955..13492,13553..13589,13671..13737))
/complement(join(12955..13492,13553..13589,13671..13737))
/note="Afa12H2.05C"
/note="Similar to Schizosaccharomyces pombe DNA 3 domain
containing protein spbc1734.05C TR:074746 (EMBL:AL031856)
(209 aa) fasta scores: E(): 2.1e-17, 41.62% id in 185 aa
prosite pattern:PS50076;dnaz domain profile:.111.351;codon
34-98
smart:SM00271;DnaJ molecular chaperone homology
domain;1.1e-08;codon 33-90
pfam:PF00226;DnaJ domain;2.9e-10;codon 34-98"
/codon_start=1
/product="hypothetical DNA 3 domain containing protein"
/protein_id="CAD37143.1"
/db_xref="GI:21627811"
/translation="MSSDEODALDALEKEASDFIKVCSIRCVSIILFLVAVLDQPG
VPSPDIQYRKKSLLIHPDKTKNPADADPAKDTLLDEKARAYLDECIADAR
RLIRREHYTVDSDELQTEFERKKEMKQATVOVLEEARRRRLKAKLOEGRKRE
EEELPARKRKELQAMENTRDERIGSRREQKQKQKKTDDKKKKMKVGLG"
/join(13982..15673,15761..15826,15869..16207,16235..16311,
16336..16414)
/gene="Afa12H2.06"
/join(13982..15673,15761..15826,15869..16207,16235..16311,
16336..16414)
/note="Similar to Schizosaccharomyces pombe probable
ATP-dependent RNA helicase spcc1093.05 TR:Q9UTP9
(EMBL:AL132839) (735 aa) fasta scores: E(): 1.5e-118,
53.67% id in 749 aa
smart:SM00490;helicase superfamily c-terminal
domain;8.5e-25;codon 305-388
smart:SM00487;DEAD-like helicases superfamily, catalytic
domain;5.6e-53;codon 66-269
pfam:PF00271;helicase conserved C-terminal
domain;3.6e-27;codon 305-388
prosite profile:PS00039;DEAD-box subfamily ATP-dependent
helicases signature Confirmed by InterPro eMotif pattern
match :8e-5;codon 198-206"
/codon_start=1
/product="probable ATP-dependent RNA helicase"
/protein_id="CAD37144.1"
/db_xref="GI:21627812"
/translation="MAPAGPRTGCHAKPQSKTLKKRGOEELSILQIVEDLDLKG
IFKFSFSLSEPTASGLASSHYRTLIDQISRALSHALKGDAITGAATGSGKTLAFL
VPEVNLVLRKQMAHEHGDGALILSTPRLEIAOIEVULKIRYRTILVIGGSLK
EOERLEGMNITLVCTPGMLQHLQOTAFYVNGMLVLDADRLDGLFOQTDAIIT
GHLPERKQTLFSAOTKQVSDLARLSIDPEYVAVHTASADSPSKLOOHYVITPLP
OKDLIMGFITSNLKSKMVLTLSSGKQVARYPYESFRHIOPTIPLMLHLGRKQGRDL
IVTRFSQSKHCVLSTVDAANGLDPAVDAYLQDPCDADDTYTHRGKATYKEGR
AVLPLDPEEGEGLMRLEQKRVPEKIRINKANKQSLIDQONMCFKQPELKYQOKA
FISYVKSYYIQDKDEIFKLKELKIDEFASGLGAPRIKIFKIDTQORNNARAA
HLSDDDTDEDEGEKSKKKEEPOVRYRMEERRQDVLAIEHYSLINDGTMOL
GSGDEDESEKGNKNVVRREKTLKILFKFGCTGLYVDDEGNPHLEVLIED
EEQKARADAKDOQAKPLAEAEKRLIADMDKTELAKKRRKKKEKRRARRELLAEA
EEETLVOLPPEYEGDDEAPRPSKRPVKTENKTEADREAEPCQDAATTPQIOTLED
LESATGLG"
/join(17274..17294,17367..17626,17679..17880,17933..17974,

CDS

18033..18181,18245..18709,18759..19248)
 /gene="AF12H2.07"
 join(17274..17294,17367..17626,17679..17880,17933..17974,
 18033..18181,18245..18709,18759..19248)
 /gene="AF12H2.07"
 /note="Similar to Yersinia pestis putative lysine-specific
 permease lypP TR:CA090139 (EMBL:AJ14147) (503 aa) fasta
 scores: E(): 1.8e-78, 42.41% id in 488 aa

Alignment Scores:

Pred. No.: 204 Length: 79718
 Score: 91.50 Matches: 44
 Percent Similarity: 36.84% Conservative: 26
 Best Local Similarity: 23.16% Mismatches: 69
 Query Match: 9.12% Indels: 51
 DB: 8 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x AF12H2 (1-79718)

QY 18 AsnAlaTfYAla-----IleGluYsgLYLysArgLeu 28
 |||:|||||
 Db 75473 AACCTCTATGACACCTCTCACTACATCGTCGCGACGCCGACACAGCGCTCGCATT 75414
 QY 29 PhePheAlaGlyLeuArgGluMet-----TyrProGlyArgAla 41
 |||:|||||
 Db 75413 CCATTGGCTTCTCTCTCGAGATGAACGCAAGTCTCATGACACTACCC----- 75363
 QY 42 LeuProSerAspAlaAsnPhetMetArgMetSerTyrGlySerIleYsgLYTArgLeu 61
 |||:|||||
 Db 75362 ---CCTTCAGACACTGATTTCTCTCTCCGACATACGCGCGCGCTTCAACTCG 75306
 QY 62 ProGlnAspGlyAlaTfYrAsnTfYrHisThrThrGlyYsgLYLeuGln 81
 |||:|||||
 Db 75305 GAGCTTCGTTCTCTCTCCAGACATACATACCGCC----- 75270
 QY 82 AspProLYSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArg----- 99
 |||:|||||
 Db 75269 ---CCCGCGTGGACTCGCTGCTTCAGCCGACGAGATCGACAGCGCTCGACATT 75213
 QY 100 ---ThrLYAsnTfYrGlyArgTfYrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSer 118
 |||:|||||
 Db 75212 ATGACAGAGAAATTGTAGCGGGTACTGAGCCGTGGAATGATTGTGTGTCGAT 75153
 QY 119 AsnAspAspIleThrGlyGlyAsnSerGlySerProValPheAspLYsAsnGlyArgLeu 138
 |||:|||||
 Db 75152 AAGACTGATCGCTCGCGCGGACGCGC-----CACGACTTCCGAGTGCAGT 75105
 QY 139 IleGlyLeuAlaPheAspGlyAsnTfYrGluAlaMetSerGlyAspIleGluPheGluPro 158
 |||:|||||
 Db 75104 CGCGGTCTCGCGCGGCGGATGTGTGGAAGATATCAG----- 75066
 QY 159 AspLeuGlnArgThrIleSerValAspIleArgTfYrValLeuPheMet----- 174
 |||:|||||
 Db 75065 ---CTGATGCTGCTTGTGCTGTGTGCTCTCATCTATCTTTCGTGGTATG 75009
 QY 175 -----IleAspLYSTfYrGlyGlnCys 181
 |||:|||||
 Db 75008 GGATCGCGCTTCCACACTTGGGGGAGATGT 74979

RESULT 7
 CDS01RG9 180688 bp DNA linear PRI 26-JUN-2001
 LOCUS Human chromosome 14 DNA sequence BAC R-63812 of library RCT-11
 DEFINITION from chromosome 14 of Homo sapiens (human), complete sequence.
 ACCESSION AL157871.5 GI:14572586
 VERSION AL157871.5 GI:14572586
 KEYWORDS HTG; HTGS; ACTIVEFIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 180688)
 AUTHORS Helling, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Wincker, P.,

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Brotier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
 Levy, M., Eckenberg, R., Bruls, T., deBardins, V., Cruaud, C.,
 Gyapay, G., Saurin, W. and Weissbach, J.
 Sequencing of the human chromosome 14
 Unpublished
 2 (bases 1 to 180688)
 Genoscope.
 Direct Submission
 Submitted (26-JUN-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
 Web : www.genoscope.cns.fr
 On Jun 27, 2001 this sequence version replaced g1:11611154.
 ----- Genome Center
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: Sequef@genoscope.cns.fr

The following BAC sequence is oriented from the r7 to the sp6 end.
 upstream BAC (overlapping the r7 end) : R-775G15
 Downstream BAC (overlapping the sp6 end) : R-362L22 (AC=AL135838)
 ----- Summary Statistics
 Assembly program: Phrap; version 2.0
 Quality coverage: 6.78x in Q20 bases; sum-of-contigs

 Overall quality chart :
 Range : Bases

0	:	:
1 - 9	:	6
10 - 19	:	83
20 - 29	:	273
30 - 39	:	1283
40 - 49	:	5543
50 - 59	:	10660
60 - 69	:	11794
70 - 79	:	22509
80 - 89	:	49844
90 - 99	:	78693

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES

source

1..180688
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone_lib="RPCT-11"
 37623..37869
 /note="matching EMBL:267128
 Rdb: RH53816
 Rdb: RH3709
 dbSTS:STS5886
 Identified using the e-PCR software (G. Schuler)"
 74837..74962
 /note="matching EMBL:G11457
 Rdb: RH1924
 dbSTS:STS19283
 Identified using the e-PCR software (G. Schuler)"
 85576..85699
 /note="matching EMBL:M77698
 Rdb: RH44091
 Rdb: RH1618
 dbSTS:STS16477
 Identified using the e-PCR software (G. Schuler)"
 86726..86933
 /note="matching EMBL:H38818
 Rdb: RH68412
 dbSTS:STS48319
 Identified using the e-PCR software (G. Schuler)"
 87142..87270
 /note="matching EMBL:T78109
 Rdb: RH25853

[illegible]

	Db	94813	AAACCTGGAGGCGCATGCACAAGAACAGGAGGGGCGCTGGCCAGGGCAAGGCTCCAGGTCC	94754
	Oy	106	TyrAlaGluAsnnglyGlnIleuHisIleAlaPheLeuSerAsnAsnAspIleThrGlycyl	125
	Db	94753	CATGCTCCTGGGTCCTCACCTTCTTGCGCTTCTTACACACCCTGACTCATG-----	94700
	Oy	126	AsnserGlySerProValAlaPheaspysAsnnglyArgIleuileGlyLeuAla-----	142
	Db	94699	-----GTACCCACCTTACCACCGAGAGGCAATGCATTGCTGGCGGCACGAGCCAC	94649
	Oy	143	PheaspGlyAsnTrpGluAlaMetserGlyAspIleGluPheGluProAspLeuInarg	162
	Db	94648	TTCCACGGGAAGTGGGAAAG-----	94628
	Oy	163	ThrIleSerValAspIleArgTyrtYrValIeuPheMetIleaspGlyTrpGlyIncysPro	182
	Db	94627	-----CGGTGGGGCCAGTGTCOA	94610
	RESULT	8		
	LOCUS	CNS01DVI		
	DEFINITION	CNS01DVI	187710 bp DNA linear PRI 04-OCT-2001	
	ACCESSION	Human chromosome 14 DNA sequence BAC R-362L22 of library RPCT-11		
	VERSION	ALI35838	from chromosome 14 of Homo sapiens (Human), complete sequence.	
	KEYWORDS	ALI35838..6 GI:15982201		
	SOURCE	Htg.		
	ORGANISM	human.		
	REFERENCE	Homo sapiens		
	AUTHORS	Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 187710) Helliö,R., Pettit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P., Brolier,P., Caticolico,L., Barbe,V., Pelletier,E., Artiguenave,F., Ley,M., Eckenberg,R., Bruls,T., debernardinis,V., Cruaud,C., Gayay,G., Saurin,W. and Weissenbach,J. Sequencing of the human chromosome 14 unpublished 2 (bases 1 to 187710) Genoscope. Direct Submission Submitted (04-OCT-2001) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr) On Oct 5, 2001 this sequence version replaced gi:13872725. ----- Genome Center Center: Genoscope / Centre National de Sequencage Center code: GS Web site: http://www.genoscope.cns.fr/ Contact: SegRef@genoscope.cns.fr ----- The following BAC sequence is oriented from the T7 to the SP6 end, Upstream BAC (overlapping the T7 end) : C-2644121 (AC-AL163974) Downstream BAC (overlapping the SP6 end) : R-63812 ----- Summary Statistics Assembly program: Phrap; version 2.0 Quality coverage: 6.92x in Q20 bases; sum-of-contigs ----- Overall quality chart : Range : bases 0 : 1 1 - 9 : 53 10 - 19 : 407 20 - 29 : 1172 30 - 39 : 5394 40 - 49 : 11391 50 - 59 : 10988 60 - 69 : 11109 70 - 79 : 20167 80 - 89 : 49419 90 - 99 : 77609 -----		

REFERENCE	AUTHORS	TITLE	JOURNAL
3	(bases 1 to 1554)	Sapats,S.I., Ashton,F., Wright,P.J. and Ignjatovic,J.	Submitted (26-MAR-1996) Sandra I. Sapats, Division of Animal Health, CSIRO, Crn Flemington Rd and Park Dr, Melbourne, VIC 3052, Australia
11	Submitted (14-JAN-1997)	Sandra I. Sapats, Division of Animal Health, CSIRO, Crn Flemington Rd and Park Dr, Melbourne, VIC 3052, Australia	Sequence update by submitter
REMARK	COMMENT	On Feb 28, 1997 this sequence version replaced g1:1515372.	
FEATURES	source	Location/Qualifiers	
		1..1554	
		/organism="Avian infectious bronchitis virus"	
		/strain="Q3/88"	
		/db_xref="taxon:11120"	
		1..1227	
		/codon_start=1	
		/product="nucleocapsid protein"	
		/protein_id="AAB48161.1"	
		/db_xref="GI:1515373"	
		/translation="MSAGKSKSDSPADILKSNVGHAPPKVSGTQGSWFOSLKEKKRTGPPTFEESGVDNSNVKQFQHGHWKRRKRSKSGKRPADAMNFEYGTGPFGLDWMGDNDVVMKKKADGTSIKGNVGRDPKPOAPLRTIEGPNPMEFIALSRGNGSSALSTRSSRPSRDSRSRORSKVITGIDRAKILIMOOKNGAISKOKANMAEKYHKRAVAAGRKIDVEFGQRKKAPNRPDDDKMIEGVADGRITAMLVPTPACLGSRVTAILOPDGLHVSFEFEYVVRDPNDSTICDAGIDGVGTRPDKEPKRSNSASDKRSAPATPKQOAKRVOKKEEERSHSEDEKEVNQLEDVDDSDIIPNKIDMGECAFDDINI"	
		PKRIDMGECAFDDINI"	
		Gaps: 7	
		Indels: 45	
		Mismatches: 44	
		Conservative: 17	
		Matches: 37	
		Length: 1554	
		Score: 90.50	
		Percent Similarity: 37.76%	
		Best Local Similarity: 25.87%	
		Identity Match: 9.02%	
		DB: 14	
		BASE COUNT	
		512 a 257 c 385 g 400 t	
		ORIGIN	
		US-10-008-355-2_COPY_522_712 (1-191) x ATU52600 (1-1554)	
		Alignment Scores:	
		Pred. No.:	
		2.24	
		Score:	
		90.50	
		Percent Similarity:	
		37.76%	
		Best Local Similarity:	
		25.87%	
		Identity Match:	
		9.02%	
		DB:	
		14	
		Gaps:	
		7	
		Indels:	
		45	
		Mismatches:	
		44	
		Conservative:	
		17	
		Matches:	
		37	
		Length:	
		1554	
		Score:	
		90.50	
		Percent Similarity:	
		37.76%	
		Best Local Similarity:	
		25.87%	
		Identity Match:	
		9.02%	
		DB:	
		14	
		Gaps:	
		7	
		Indels:	
		45	
		Mismatches:	
		44	
		Conservative:	
		17	
		Matches:	
		37	
		Length:	
		1554	
		Score:	
		90.50	
		Percent Similarity:	
		37.76%	
		Best Local Similarity:	
		25.87%	
		Identity Match:	
		9.02%	
		DB:	
		14	
		Gaps:	
		7	
		Indels:	
		45	
		Mismatches:	
		44	
		Conservative:	
		17	
		Matches:	
		37	
		Length:	
		1554	
		Score:	
		90.50	
		Percent Similarity:	
		37.76%	
		Best Local Similarity:	
		25.87%	
		Identity Match:	
		9.02%	
		DB:	
		14	
		Gaps:	
		7	
		Indels:	
		45	
		Mismatches:	
		44	
		Conservative:	
		17	
		Matches:	
		37	
		Length:	
		1554	
		Score:	
		90.50	
		Percent Similarity:	
		37.76%	
		Best Local Similarity:	
		25.87%	
		Identity Match:	
		9.02%	
		DB:	
		14	
		Gaps:	
		7	
		Indels:	
		45	
		Mismatches:	
		44	
		Conservative:	

[illegible]


```

Db 1696 -----GGGCTCTGTGGACGGCATATTGCCAGCAGCAAGCTGAGCGCAATGCT 1746
QY 78 LeuGluLyGlnAspProLySerSerSgluPheAlaValGlnGluAsnIleuAspLeu 97
Db 1747 TCTGATAAAGAGCGCGGCAACAGCGGGAATTTACTGTG-----1785
QY 98 PheArgThrLyAsnTyrGlyArgTyrAlaGluAsnGlyGlnIleuHisIleAlaPheLeu 117
Db 1786 -----AATTTGGCGAGAAAAAATTACCGCAGCCTTA-----1818
QY 118 SerAsnAsnAspIleThrGlyGlnAsnSerGlySerProValPheAspLySasnglyArg 137
Db 1819 -----ACCGCTGAAAAACAGCAGCAGGCAACCTTTACATTCATGTGTAG 1863
QY 138 LeuIleGlyLeuAlaPheAspLySasnglyArgTyrGluAlaMetSerGlyAspIleGluPheGlu 157
Db 1864 ATTTAGAGGCAACGCTTTTCCGCTAGC-----GCAAAAAGCTGCAATTAAGTTTGTAT 1917
QY 158 ProAspLeuGlnArgThrIleSerValAspIleArgTyrVal 171
Db 1918 CTCGATCAAAAAAATACACCCGCGCCTTAAGCATATATC 1959

RESULT 11
LOCUS NMTBP2163 2163 bp DNA linear BCT 31-JAN-1996
DEFINITION N.meningitidis DNA for tbp2 gene (strain B2163).
ACCESSION 250731
VERSION 250731.1 GI:1177568
KEYWORDS Tbp2; Transferin-binding protein 2.
SOURCE Neisseria meningitidis.
ORGANISM Neisseria meningitidis.
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria
REFERENCE 1 (bases 1 to 2163)
AUTHORS Legrain,M., Pindel,A., Villevet,D., Quentin-Millet,M. and
JACOBS,E.
TITLE Molecular characterization of hybrid transferrin-binding protein
2's from Neisseria meningitidis
JOURNAL Unpublished
AUTHORS Legrain,M.
TITLE Direct Submission
JOURNAL Submitted (03-AUG-1995) Legrain M., Transgene, Molecular Genetics,
11, rue de Molshelm, Strasbourg, Bas-Rhin, France, 67000
FEATURES
source location/Qualifiers
1..2163
/organism="Neisseria meningitidis"
/strain="B2163"
/db_xref="taxon:487"
1..2136
/gene="tbp2"
/gene="tbp2"
1..2136
/codon_start=1
/transl_table=1
/product="transferrin-binding protein 2 precursor"
/protein_id="CA90598.1"
/db_xref="GI:1177569"
/db_xref="SPTREMBL:O51284"
/translation="MNNPLVNOAMVLFVLSACLGSGGSHLDSDVTEAPRPARY
ODVSEKPOADGOGGEFAMLRKRNRHPOKKEKVELNPDMDETGTPSKPNLPE
ROQSYIDKVTGDSNITYSPILQSNHONGSTNGCANQPKREVDYNNKRYVSGWF
YKHAISREPKIKRSGDDGITFTHGKDPSSQDPTSEKVTIKGVWHVTDTEKQKF
NDILETSKGGGRYSFGSDGETTSNRDNLNKHGGYGTSLMEVDFGSKLTGK
LIRNRVNTATNDKRYTYOYSLDAQITGNRNGKAIATDKPDGTGTLHPVSDSS
LSGFGFRGELGFRFLSDKVAVVGSAKTKDKENGAVAGSTDAASAGAGTS
SENSKLTIVLDAVELGDKREYOKLDNFNSAOLVVDGIMPLPETSSEGNMONG
TNGGATFRKEDHPESDKDAQCTONGAQTANTGQDNTNGKTKTYEVECCSNL
LYKGMILTRKNSKAMQAGESSQADAKTBOVGSMFLDGETIDEREKEIIPSEQNIYRG
SWYGHIASTWSGASNADEKGNRAEFTVNEKEKTIQTGLTAEINQDEATFTDGIIEG
NGFSGTARTAEIGFDLDQKNTTRPKAVITDAKVOGFGYGPRAEELGGMFAVOGDQOT
ENTVVASGNGNSASATVVFCAKROKRPVO"

```

```

sig_peptide 1..60
mat_peptide /gene="tbp2"
/gene="tbp2"
/product="Transferrin-binding protein 2"
BASE COUNT 706 a 480 c 529 g 448 t
ORIGIN
Alignment Scores:
Pred. No.: 3.79 Length: 2163
Score: 90.00 Matches: 38
Percent Similarity: 37.93% Conservative: 28
Best Local Similarity: 21.84% Mismatches: 66
Query Match: 8.97% Indels: 42
Gaps: 7
US-10-008-355-2_COPY_522_712 (1-191) x NMTBP2163 (1-2163)
QY 2 LysSerValIleAlaAlaIleAlaAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAla 21
Db 1552 AAAAAGCAGACAGTCCCGCATGCGAGCAGGAGAAAGCAGTACAGTCACTGCTAAAG 1611
QY 22 IleGluLyGlyLysArgLeuPhePheAlaGlyLeuArgGluMetArgProGlyArgAla 41
Db 1612 GAACAAGTTGACAAAGTATGTTCTCCCAAGGC-----GAGCGCAGCATGAAAAAGAG 1665
QY 42 LeuProSerAspAlaAsnSerMetArgMetSerTyrGlySerIleLySglTyrGlu 61
Db 1666 ATTCAGAGCAGCAAAACATCGTTATCGG-----1695
QY 62 ProGlnAspGlyAlaTyrPyrAsnTyr-----HisThrThrGlyLysGly--Val 77
Db 1696 -----GCGCTGTGTACGGCGCATATTGCCAGCAGCAAGCTGAGCGCAATGCT 1746
QY 78 LeuGluLyGlnAspProLySerSerSgluPheAlaValGlnGluAsnIleuAspLeu 97
Db 1747 TCTGATAAAGAGCGCGCAACAGCGCGGAATTTACTGTG-----1785
QY 98 PheArgThrLyAsnTyrGlyArgTyrAlaGluAsnGlyGlnIleuHisIleAlaPheLeu 117
Db 1786 -----AATTTGGCGAGAAAAAATTAACCGCAGCCTTA-----1818
QY 118 SerAsnAsnAspIleThrGlyGlnAsnSerGlySerProValPheAspLySasnglyArg 137
Db 1819 -----ACCGCTGAAAAACAGCAGCAGGCAACCTTTACATTCATGTGTAG 1863
QY 138 LeuIleGlyLeuAlaPheAspLySasnglyArgTyrGluAlaMetSerGlyAspIleGluPheGlu 157
Db 1864 ATTTAGAGGCAACGCTTTTCCGCTAGC-----GCAAAAAGCTGCAATTAAGTTTGTAT 1917
QY 158 ProAspLeuGlnArgThrIleSerValAspIleArgTyrVal 171
Db 1918 CTCGATCAAAAAAATACACCCGCGCCTTAAGCATATATC 1959

RESULT 12
LOCUS AK023006 2794 bp mRNA linear PRI 01-AUG-2002
DEFINITION AK023006 AK023006
to Homo sapiens cDNA FLJ12944 f1s, clone NTZRP2005168, highly similar
to Homo sapiens mRNA for E1B-55kDa-associated protein.
ACCESSION AK023006
VERSION AK023006.1 GI:10434722
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens testicular carcinoma cell line:NT2 cDNA to mRNA,
clone_11b:NT2RP2 clone:NT2RP2005168.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1
AUTHORS Isoqai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Takahashi-Fujii,A., Hara,H.,
Tanase,T., Nomura,Y., Togiya,S., Komei,F., Hara,R., Takeuchi,K.,
Arita,M., Nabekura,T., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J.,
Wakamatsu,A., Nakamura,Y., Nageharu,K., Masuo,Y. and Oshima,A.

```

[illegible]

Db	998	TGCGACGTGAGCTCACTTC	918
RESULT 13			
LOCUS	BC021506		
DEFINITION	BC021506	2908 bp	mRNA linear ROD 07-AUG-2002
ACCESSION	BC021506		
VERSION	BC021506.1		
KEYWORDS	GI:18204831		
SOURCE	Mus musculus.		
ORGANISM	house mouse.		
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2908)		
TITLE	Strausberg, R.		
JOURNAL	Direct Submission		
REMARK	Submitted (14-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: gcgaps-remail.nih.gov Tissue Procurement: Jeffrey Green M.D. DNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nhgri.nih.gov Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Green, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karins, E., Kiong, P., Lalic, P., Legaspi, R., Maduro, Q. L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Statulic, S., Thomas, P. J., Touchman, J.W., Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://Image.Llnl.gov Series: IRAC Plate: 55 Row: 1 Column: 3 This clone was selected for full length sequencing because it passed the following selection criteria: Genomescan gene prediction.		
SOURCE	Location/Qualifiers		
	1. .2908		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	/map="FVB/N"		
	/clone="MGC:36621 IMAGE:5347099"		
	/tissue_type="Mammary tumor, C3(1)-Tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."		
	/clone_11b="NCI_CGAP_Mam6"		
	/lab_host="DH10b"		
	/note="Vector: pCMV-Sport6"		
	110. .2689		
	/codon_start=1		
	/product="Similar to E1B-55kDa-associated protein 5"		
	/protein_id="AAH21506.1"		
	/db_xref="GI:18204832"		
	/translation="MDYRRARVLEBELQRRDLDRGLKALAEFLAALAEPEEDERELEADDDPGLPGHNEEVEETVGGSELESTGAPPPGLQHPPEEGVSGDDHYWDNTRONQDEETPVYIKQENESSYDRPLDMEPOOVYHPELKTMEKQAPSPFLPYASOLDTRFQONRRKRPPEENRGSGYFEHREDRGRSPQPAEDEDDEFDITLVAIDRYNCLDHPKAFDARRSSGYPITIGFAVILWSGAASGVBRGVCPEMKINETSIVHLSTEPHPYVIRIGMSKDCSTQIGEPSPGKSKTSNPFENYGDKFAENVYIGGFADEPCGAVELSPTRKMKMKIARIQKALGGCALYPHYVKKCAVEFNQGAPEKSVLDEGFTHLPUSERIKRTIGPKSADEILMMVGLPAGVTTVAIKHASPSPKYNIGINAIMDKRMVGLRQRYNAGVMDVLIQAOCLNRLIOIARKKRNITDQTNVYSGQRMRMPFEFGSKRAIYICPTDELDKRTYKRTDDECKDVPDHAIVLEMKANF		

FEATURES	SOURCE
53	---serTtYcGlySerTleYsclYyrgLuprOgInAspGlyAlaTrpTyrAsnTyrHis 71
927	GACCTCGCAGCAGCCAGCTAGGCAAGAGCCCTTCTCC-----TATGGCTATGGA 977
72	TThrTgGlyLysGlyYValLeuGLuLysGlnAspProLysSerAspGluPheAlaValGln 91
978	GGCACTGGGAAC-----AACTCCACCAAT-----1001
92	GluAsnLleLeuAspLeuPheArgTrhTyrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110
1002	-----AGCCGGATTGAAAACTACGCGAGACAAAGTTGACAGAG-----1037
111	GlnLeuHisLleAlaPheLeuSerAsnAsnAspLleThrGly-----124
1038	-----AACGATGATGATTGGCTGCTTTGGCGGATTTTGA 1070
125	---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
1071	TGTGGAATATGACGTGAGACTGCTTTTACCAAGATGGAAGATGATGGATGGCATTCCTTC 1130
144	AspGlyAsnTrpGluAlaMetSerGlyAspLleGluPheGluProAspLeuGlnArgThr 163
1131	CGAATCCAGAAAGAGCCCTTGGGGGTCAGGCGCTCTATCTCATGCTCTGTGTAAGAAT 1190
164	IleSerValAspIleArgTyr 170
1191	TGGCGCAGTGGAGTTCAACTTC 1211
RESULT 15	
LOCUS	BC009988 3048 bp mRNA linear PRI 12-JUL-2001
DEFINITION	Homo sapiens, clone MGC:16706 IMAGE:4127873, mRNA, complete cds.
ACCESSION	BC009988
VERSION	BC009988.1 GI:1460312
KEYWORDS	MGC.
SOURCE	Homo sapiens.
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 3048)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (02-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgabbs-remail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/nisc.mgc@nhgri.nih.gov Contact: Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M., Bejarlun,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S., Dietrich,N.B., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R., Lim,M., Maduro,Q.L., Mastello,C., Mastrian,S.D., McLooskey,J.C., McOmwall,J., Pearson,R., Snyder,B., Stentitpop,S., Thomas,P.J., Tlionson,E.E., Touchman,J.W., Tsurgon,C., Vogt,J.L., Walker,M.A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: <http://image.lnl.gov>
Series: IRAL Plate: 26 Row: h Column: 8.
Location/Qualifiers
1. 3048
/organism="Homo sapiens"
/db_xref="taxon:9606"

[illegible]

Search completed: December 21, 2002, 02:43:11
Job time : 3078 secs

FEATURES	Location/Qualifiers
source	1. .3048
	/organism="Homo sapiens"
	/db_xref="taxon:9606"

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 07:26:30 : Search time 48.3158 Seconds
(without alignments)
38.381 Million cell updates/sec

Title: US-10-008-355-26
Perfect score: 48
Sequence: 1 TCGNSGSPV 9

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPPREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_plant:*
10: sp_rodent:*
11: sp_virus:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	316	2	Q9FBG1
2	48	100.0	342	16	Q99V45
3	48	100.0	357	2	Q04186
4	48	100.0	716	16	Q9PC94
5	43	89.6	217	2	Q9AJX0
6	42	87.5	169	16	Q8YV87
7	41	85.4	284	2	Q47809
8	40	83.3	315	16	Q9ZDX8
9	40	83.3	353	11	Q62313
10	40	83.3	363	11	Q62314
11	40	83.3	497	16	Q9ZJ35
12	39	81.2	396	12	Q8VAD4
13	39	81.2	499	10	Q9LKT0
14	39	81.2	996	3	Q9P7S1
15	39	81.2	997	3	Q74325
16	38	79.2	253	2	P96151

17	38	79.2	289	16	Q98M08	Q98M08 rhizobium 1
18	38	79.2	394	12	Q8OR15	Q8OR15 apple stem
19	38	79.2	398	11	Q8VE75	Q8VE75 mus musculus
20	38	79.2	476	16	Q9KLD4	Q9KLD4 vibrio chol
21	38	79.2	593	10	Q9FY14	Q9FY14 arabidopsis
22	38	79.2	613	11	Q9D2L9	Q9D2L9 mus musculus
23	38	79.2	627	10	Q9SRP2	Q9SRP2 arabidopsis
24	38	79.2	1097	10	Q8RY22	Q8RY22 arabidopsis
25	37	77.1	235	2	Q9KH51	Q9KH51 staphylococ
26	37	77.1	235	2	Q9FD08	Q9FD08 staphylococ
27	37	77.1	235	16	Q99760	Q99760 staphylococ
28	37	77.1	239	2	Q9KH49	Q9KH49 staphylococ
29	37	77.1	239	16	Q53782	Q53782 staphylococ
30	37	77.1	240	2	Q9KH50	Q9KH50 staphylococ
31	37	77.1	240	16	Q53781	Q53781 staphylococ
32	37	77.1	284	16	Q93JK9	Q93JK9 streptomyce
33	37	77.1	370	16	Q9KZV9	Q9KZV9 streptomyce
34	37	77.1	449	16	Q83887	Q83887 treponema p
35	37	77.1	461	4	Q9Y414	Q9Y414 homo sapien
36	37	77.1	590	16	Q8Y3Y8	Q8Y3Y8 listeria mo
37	37	77.1	817	5	Q93560	Q93560 caenorhabdi
38	37	77.1	857	5	Q8T070	Q8T070 drosophila
39	37	77.1	1008	16	Q98KC9	Q98KC9 rhizobium 1
40	37	77.1	1117	4	Q9H5J5	Q9H5J5 homo sapien
41	37	77.1	1254	4	Q9NS14	Q9NS14 homo sapien
42	37	77.1	1306	16	Q53775	Q53775 mycobacteri
43	37	77.1	1377	4	Q9UPN5	Q9UPN5 homo sapien
44	37	77.1	1844	5	Q22579	Q22579 caenorhabdi
45	37	77.1	1971	4	Q9NTT5	Q9NTT5 homo sapien

ALIGNMENTS

RESULT 1
Q9FBG1 PRELIMINARY: PRT: 316 AA.
AC Q9FBG1:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Glutamy1 endopeptidase.
GN PROM.
OS Staphylococcus warneri.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1292;
RN [1]
RN SEQUENCE FROM N.A.
RA Submitted (Aug-2000) to the EMBL/Genbank/DBJ databases.
RL Kakiyama M.;
RN [2]
RN SEQUENCE FROM N.A.
RA Kodaira K.I.;
RT "Characterization of the gene encoding glutamy1 endopeptidase of
RT Staphylococcus warneri M.",
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ293885; CAC06168.1; -.
DR MEROPS: S01.269; -.
DR InterPro: IPR001254; Ser_protease_Try.
DR InterPro: IPR00126; Ser_proteas_V8.
DR Pfam: PF00089; trypsin.1.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00020; TRYP-Spc.1.
DR PROSITE: PS0240; TRYP-SIN_DOM.1.
DR PROSITE: PS00673; V8_SER.1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 316 AA; 34296 MW; 4E97A5A111DB40 CRC64;
Query Match 100.0%; Score 48; DB 2; Length 316;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
 |||||||
 DB 231 TCGNSGSPV 239

RESULT 2

O99V45 PRELIMINARY: PRT: 342 AA.

AC O99V45;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE Serine protease, V8 protease, glutamyl endopeptidase.
 GN SPPA OR SAV1048 OR SA0901.
 OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 Staphylococcus.
 NCBI_TaxID=158878, 158879;
 [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
 RX MEDLINE=21311952, PubMed=11418146;
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
 Karamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 Mizutani-U Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,
 Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RA "Whole genome sequencing of methicillin-resistant Staphylococcus
 RT aureus.";
 RL Lancet 357:1225-1240(2001).
 CC -1 SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AP003361; BAB57210.1; -;
 DR EMBL: AP003132; BAB42146.1; -;
 DR MEROPS: S01.269; -;
 DR InterPro: IPR001254; Ser_protease_TRY.
 DR InterPro: IPR000126; Ser_proteas_V8.
 DR Pfam: PF00089; trypsin_1.
 DR PRINTS: PR00839; VBPROTEASE.
 DR SMART: SM00020; TRY_SPC_1.
 DR PROSITE: PS50240; TRYPSIN_DOM_1.
 DR PROSITE: PS00672; V8_HIS_1.
 DR PROSITE: PS00673; V8_SER_1.
 KW Hydrolyase: Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 342 AA; 36977 MW; 5ABF42DC01C4B24 CRC64;

Query Match 100.0%; Score 48; DB 16; Length 342;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
 |||||||
 DB 233 TCGNSGSPV 241

RESULT 3

O04186 PRELIMINARY: PRT: 357 AA.

AC O04186;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE Glutamic acid specific protease prepropeptide (EC 3.4.21.19).
 OS Staphylococcus aureus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 Staphylococcus.
 NCBI_TaxID=1280;
 [1]
 RP SEQUENCE FROM N.A.
 RA Yoshikawa K., Tsuzuki H., Fujiwara T., Nakamura E., Iwamoto H.,
 RA Matsumoto K., Shin M., Yoshida N., Teraoka H.;

RT "Purification, characterization and gene cloning of a novel glutamic
 RT acid-specific endopeptidase from staphylococcus aureus atcc 12600.";
 RL Biochim. Biophys. Acta 1121:221-228(1991).
 DR EMBL: D00730; BAA00630.1; -;
 DR MEROPS: S01.269; -;

DR InterPro: IPR001254; Ser_protease_TRY.
 DR InterPro: IPR000126; Ser_proteas_V8.
 DR Pfam: PF00089; trypsin_1.
 DR PRINTS: PR00839; VBPROTEASE.
 DR PROSITE: PS50240; TRYPSIN_DOM_1.
 DR PROSITE: PS00672; V8_HIS_1.
 DR PROSITE: PS00673; V8_SER_1.
 KW Hydrolyase; Protease; Serine protease.
 FT CHAIN 69 357
 SEQUENCE 357 AA; 38651 MW; 58AA9A4E371E2577 CRC64;

Query Match 100.0%; Score 48; DB 2; Length 357;
 Best Local Similarity 100.0%; Pred. No. 1.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
 |||||||
 DB 233 TCGNSGSPV 241

RESULT 4

O9PC94 PRELIMINARY: PRT: 716 AA.

AC O9PC94;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DE Hypothetical protein Xf1887.
 GN Xf1887.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 NCBI_TaxID=2371;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.E., Bata G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Farga J.S., Franca S.C., Franco M.C., Frohne M., Furian L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madelira A.M.B.N., Madelira H.M.F., Martino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteloro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Neto L.E.S.,
 RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmeri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
 RA Quaglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidants J., Seubai J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL: AE004008; AAF84593.1; -;
 DR MEROPS: S46.001; -;
 KW Hypothetical protein: Complete proteome.

SO SEQUENCE 716 AA; 79375 MW; EAF086EE2315BBD6C CRC64;

Query Match 100.0%; Score 48; DB 16; Length 716;

Best Local Similarity 100.0%; Pred. No. 2.9; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9

DB 647 TGGNSGSPV 655

RESULT 5

O9AJX0 PRELIMINARY; PRT; 217 AA.

AC O9AJX0; 01-JUN-2001 (TREMBLrel. 17, Created)

DT 01-JUN-2001 (TREMBLrel. 21, Last sequence update)

DE Extracellular serine proteinase precursor (Fragment).

GN ESP.

OS Staphylococcus epidermidis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OX NCBI_TaxID=1282;

RP SEQUENCE FROM N.A.

RC STRAIN=6746;

RA Dubin G.;

RT "Staphylococcus epidermidis extracellular serine proteinase.";

RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.

DR EMBL; AJ305145; CAC27157.1;

DR HSSP; P09331; 1EXF.

DR MEROPS; S01.269; .

DR InterPro; IPR001254; Ser_protease_Try.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00839; V8PROTEASE.

DR PROSITE; PS50240; TRYPSIN_DOM; 1.

DR PROSITE; PS00673; V8_SER; 1.

KT Hydrolase; Serine protease; Signal.

FT SIGNAL

FT CHAIN

FT SEQUENCE

Query Match 89.6%; Score 43; DB 2; Length 217;

Best Local Similarity 100.0%; Pred. No. 6.7;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGNSSGSPV 9

DB 167 GGNSSGSPV 174

RESULT 6

O8YM87 PRELIMINARY; PRT; 169 AA.

AC O8YM87; 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)

DE Hypothetical protein ALR5049.

GN ALR5049.

OS Anabaena sp. (strain PCC 7120).

OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.

OX NCBI_TaxID=103690;

RP SEQUENCE FROM N.A.

RA MEDLINE-21595285; PubMed-11759840;

RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,

RA Matsumoto M., Matsuno A., Muraki A., Nakazawa M., Yamada M.,

RA Kishida Y., Kohara M., Sugimoto M., Takazawa M.,

RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M.,

RA Yasuda M., Tabata S.;

RT "Complete genomic sequence of the filamentous nitrogen-fixing

RT cyanobacterium Anabaena sp. strain PCC 7120."

RL DNA Res. 8:205-213(2001).

DR EMBL; AF003598; BAB76748.1;

KW Hypothetical protein; Complete proteome.

SO SEQUENCE 169 AA; 17427 MW; 0A610CC04EAC48A CRC64;

OY 1 TGGNSGSPV 9

DB 109 TGGNSASPV 117

Query Match 87.5%; Score 42; DB 16; Length 169;

Best Local Similarity 88.9%; Pred. No. 7.8;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9

DB 109 TGGNSASPV 117

Query Match 85.4%; Score 41; DB 2; Length 284;

Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9

DB 224 TGGSGSPV 232

Query Match 85.4%; Score 41; DB 2; Length 284;

Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9

DB 224 TGGSGSPV 232

Query Match 85.4%; Score 41; DB 2; Length 284;

Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9

DB 224 TGGSGSPV 232

Query Match 85.4%; Score 41; DB 2; Length 284;

Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9

DB 224 TGGSGSPV 232

Query Match 85.4%; Score 41; DB 2; Length 284;

Best Local Similarity 77.8%; Pred. No. 20;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGGNSGSPV 9

DB 224 TGGSGSPV 232

RX MEDLINE-99039499; PubMed-9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichteritz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 mitochondria."
 CC Nature 396:133-140(1998).
 RL -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AJ235270; CAI14652.1; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 KM Hydrolase; Serine protease; Complete proteome.
 SQ SEQUENCE 315 AA; 35033 MW; 2D91AD54FBE9A1 CRC64;

Query Match 83.3%; Score 40; DB 16; Length 315;
 Best Local Similarity 88.9%; Pred. No. 33;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
 ||| |||||
 Db 188 TCGASGSPV 196

RESULT 9

O62313 PRELIMINARY; PRT; 353 AA.

AC O62313; 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Trans-golgi network integral membrane protein TGN38A precursor (Trans-

golgi network protein 1) (TGN38 homolog).

GN TTGN1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.

RC STRAIN-ICR; TISSUE-BRAIN;

RX MEDLINE-95301533; PubMed-7540170;

RA Kasai K., Takahashi S., Murakami K.,

RT "Strain-specific presence of two TGN38 isoforms and absence of TGN41

in mouse."

RL J. Biol. Chem. 270:14471-14476(1995).

RN [2]

RP SEQUENCE FROM N.A.

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND

FROM TRANS-GOLGI NETWORK.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-

GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL

SURFACE RETURNING VIA ENDOSOMES (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.

CC -1- MISCELLANEOUS: ALSO FOUND IN STRAINS BALB/C, C57BL/6 AND DBA/2.

DR EMBL: D50031; BAA08757.1; -.

DR EMBL: BC009143; AA09143.1; -.

DR MGD: MGI:105080; Tgn1.

KM Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.

FT SIGNAL 1 17

FT CHAIN 18 353

FT TRANSMEM 18 298

FT DOMAIN 299 319

FT DOMAIN 320 353

FT DOMAIN 346 349

FT DOMAIN 131 178

FT REPEAT 131 138 1.
 FT REPEAT 139 146 2.
 FT REPEAT 147 154 3.
 FT REPEAT 155 162 4.
 FT REPEAT 163 170 5.
 FT REPEAT 171 178 6.
 FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 353 AA; 37848 MW; 95C340C2F4A21EB3 CRC64;

Query Match 83.3%; Score 40; DB 11; Length 353;
 Best Local Similarity 87.5%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 8
 |||||
 Db 138 TCGNSGSPV 145

RESULT 10

O62314 PRELIMINARY; PRT; 363 AA.

AC O62314; 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE Trans-golgi network integral membrane protein TGN38B precursor (Trans-

golgi network protein 2) (TGN38 homolog).

GN TGNL2 OR TTGN2.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ICR; TISSUE-BRAIN;

RX MEDLINE-95301533; PubMed-7540170;

RA Kasai K., Takahashi S., Murakami K.,

RT "Strain-specific presence of two TGN38 isoforms and absence of TGN41

in mouse."

RL J. Biol. Chem. 270:14471-14476(1995).

CC -1- FUNCTION: MAY BE INVOLVED IN REGULATING MEMBRANE TRAFFIC TO AND

FROM TRANS-GOLGI NETWORK.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. PRIMARILY IN TRANS-

GOLGI NETWORK. CYCLES BETWEEN THE TRANS-GOLGI NETWORK AND THE CELL

SURFACE RETURNING VIA ENDOSOMES (BY SIMILARITY).

CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.

CC -1- MISCELLANEOUS: NOT FOUND IN STRAINS BALB/C, C57BL/6 AND DBA/2.

DR EMBL: D50032; BAA08758.1; -.

DR MGD: MGI:105079; Tgnl2.

KM Signal; Transmembrane; Glycoprotein; Repeat; Golgi stack.

FT SIGNAL 1 17

FT CHAIN 18 363

FT TRANSMEM 18 308

FT DOMAIN 309 329

FT DOMAIN 330 363

FT SITE 356 359

FT DOMAIN 133 188

FT REPEAT 133 140

FT REPEAT 141 148

FT REPEAT 149 156

FT REPEAT 147 164

FT REPEAT 165 172

FT REPEAT 173 180

FT REPEAT 181 188

FT CARBOHYD 303 303

FT SEQUENCE 363 AA; 38821 MW; 2826FA9E958C5C27 CRC64;

Query Match 83.3%; Score 40; DB 11; Length 363;
 Best Local Similarity 87.5%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSP 8
|||||||
DB 140 TCGNSGSP 147

RESULT 11

O92J35 PRELIMINARY; PRT; 497 AA.
AC O92J35;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Heat shock protease.
GN RC0234.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiase; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MALISH 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audifren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
DR EMBL; AF008590; AL02772.1; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Ser-protease_Try.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PROSITE: PS50106; PDZ; 1.
KW Protease; Complete Proteome.
SQ SEQUENCE 497 AA; 55517 MW; 4EB5872B552EFC6C CRC64;

Query Match 83.3%; Score 40; DB 16; Length 497;
Best Local Similarity 88.9%; Pred. No. 53;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TCGNSGSP 9
|||||||
DB 188 TCGNSGSP 196

RESULT 12

O8VA04 PRELIMINARY; PRT; 396 AA.
AC O8VA04;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Coat protein.
OS Apple stem pitting virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Foveavirus.
OX NCBI_TaxID=35350;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WT32;
RA Komorowska B., Malinowski T.;
RT "Diversity of the coat protein gene sequence of several ASPV
RT isolates.";
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF438521; AL32457.1; -.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR000052; PltVir-coat.
DR Pfam: PF00286; virus_P-coat; 1.
DR PRINTS: PR00232; POTXCARCOAT.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00418; POREX_CARLAVIRUS_COAT; UNKNOWN_1.
SQ SEQUENCE 396 AA; 42574 MW; E9237780EBBC333 CRC64;

Query Match 81.2%; Score 39; DB 12; Length 396;

Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GCGNSGSP 8
|||||||
DB 171 GCGNSGSP 177

RESULT 13

O9LK70 PRELIMINARY; PRT; 499 AA.
AC O9LK70;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Similarity to serine protease.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-COLUMBIA;
RX MEDLINE=20363099; PubMed=10907853;
RA Nakamura Y.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,
RT TAC and BAC clones.";
RL DNA Res. 7:217-221(2000).
DR EMBL; AF000373; BAB01154.1; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser-protease_Try.
DR InterPro: IPR00126; Ser-proteas_V8.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolyase; Protease; Serine protease.
SQ SEQUENCE 499 AA; 56188 MW; 016D4BA02CA69BE1 CRC64;

Query Match 81.2%; Score 39; DB 10; Length 499;
Best Local Similarity 87.5%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 GCGNSGSP 9
|||||||
DB 219 GCGNSGSP 226

RESULT 14

O9P7S1 PRELIMINARY; PRT; 996 AA.
AC O9P7S1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Hypothetical signalling-associated PDZ domain containing protein.
GN SPAC3363.12C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-972H-;
 RA Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL139854; CAB72237.1; -.
 DR InterPro: IPR002114; HPr_SerP_site.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR Pfam: PF00595; PDZ; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 3.
 DR PROSITE: PS00589; PTS_HPR_SER; UNKNOWN_1.
 SQ SEQUENCE 996 AA; 110407 MW; A008B51746B05AC3 CRC64;

Query Match 81.2%; Score 39; DB 3; Length 996;
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
 :||:|||||
 DB 212 SGGSSGSPV 220

RESULT 15

O74325 PRELIMINARY; PRT; 997 AA.
 AC O74325;
 DT 01-NOV-1998 (TREMblrel. 08, Created)
 DT 01-NOV-1998 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical 111.3 Kda protein C1685.05 in chromosome II.
 GN SPBC1685.05.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-972;
 RA Wood V., Rajandream M.A., Barrell B.G., Hiltbert H., Duisterhoeft A.;
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: TO YEAST YNL123W.
 DR EMBL: AL031154; CAA20053.1; -.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00089; trypsin; 1.
 DR SMART: SM00228; PDZ; 3.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 997 AA; 111292 MW; D32D7E3CCA877A0 CRC64;

Query Match 81.2%; Score 39; DB 3; Length 997;
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCGNSGSPV 9
 :||:|||||
 DB 227 SGGSSGSPV 235

Search completed: December 20, 2002, 12:10:22
 Job time : 52.3158 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 20, 2002, 17:30:44 ; Search time 51 Seconds
(without alignments)
1148.535 Million cell updates/sec

Title: US-10-008-355-2_COPY_522_712

Perfect score: 1003

Sequence: 1 SKSVIAAARAIOADAMANAY.....LFMDKWKQCPRLIOELKLI 191

Scoring table:
BLOSUMP62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framed.p2n.model -DEV=xjh
-O/cgnt2_1/USPRO.spool/US10008355/runat_17122002_112703_16908/app_query.fasta_1.327
-DB=Issued_Patents_NA -QPM=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosump62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10008355 @CGN1.1.17 @runat_17122002_112703_16908 -NCPU=6 -ICPU=3
-NO_XLPPY -NO_MMAP -LARGEOURRY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-MARKTIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database :

- 1: Issued Patents_NA.*
- 2: /cgnt2_6/ptodata/1/ina/5A.COMB.seq.*
- 3: /cgnt2_6/ptodata/1/ina/5B.COMB.seq.*
- 4: /cgnt2_6/ptodata/1/ina/6A.COMB.seq.*
- 5: /cgnt2_6/ptodata/1/ina/6B.COMB.seq.*
- 6: /cgnt2_6/ptodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	97.4	97.4	1974	US-09-221-017B-726	Sequence 726, App
2	427.5	42.6	2384	US-09-221-017B-1045	Sequence 1045, Ap
3	89.5	8.9	10862	US-09-058-411-1	Sequence 1, Appli
4	83	8.3	1083	US-09-134-001C-1520	Sequence 1520, Ap
5	81	8.1	843	US-09-660-587-3	Sequence 3, Appli
6	81	8.1	849	US-09-261-358A-3	Sequence 3, Appli
7	77.5	7.7	841	US-08-990-823-39	Sequence 39, Appli
8	77.5	7.7	4403765	US-09-103-840A-2	Sequence 2, Appli
9	77.5	7.7	4411529	US-09-103-840A-1	Sequence 1, Appli
10	75	7.5	11672	US-09-441-340-2	Sequence 2, Appli
11	74.5	7.4	11459	US-09-454-721A-3	Sequence 3, Appli
12	74	7.4	2989	US-08-433-522A-9	Sequence 9, Appli

13	74	7.4	2989	3	US-09-135-166-9	Sequence 9, Appli
14	74	7.4	2989	4	US-08-942-046-9	Sequence 9, Appli
15	73.5	7.3	2934	4	US-09-206-942-52	Sequence 52, Appli
16	73.5	7.3	2952	4	US-09-206-942-50	Sequence 50, Appli
17	73.5	7.3	3270	4	US-08-637-732A-1	Sequence 1, Appli
18	73.5	7.3	10718	3	US-08-325-426B-1	Sequence 1, Appli
19	73	7.3	5741	1	US-07-706-699-4	Sequence 4, Appli
20	73	7.3	5741	1	US-07-998-931-4	Sequence 4, Appli
21	73	7.3	11517	1	US-07-920-281C-1	Sequence 1, Appli
22	73	7.3	11517	4	US-08-466-277-1	Sequence 1, Appli
23	72.5	7.2	842	3	US-08-733-230-3	Sequence 3, Appli
24	72.5	7.2	842	4	US-08-953-326-3	Sequence 3, Appli
25	72.5	7.2	2230	3	US-08-448-194-7	Sequence 7, Appli
26	72.5	7.2	2230	4	US-08-867-921-7	Sequence 7, Appli
27	72	7.2	506	1	US-08-469-802B-7	Sequence 7, Appli
28	72	7.2	506	2	US-08-267-803B-7	Sequence 7, Appli
29	72	7.2	2174	4	US-07-613-444-1	Sequence 1, Appli
30	72	7.2	2974	3	US-08-433-522A-7	Sequence 7, Appli
31	72	7.2	2974	4	US-09-135-166-7	Sequence 7, Appli
32	72	7.2	2974	4	US-08-942-046-7	Sequence 7, Appli
33	72	7.2	10660	2	US-08-267-803B-8	Sequence 8, Appli
34	72	7.2	10660	4	US-09-041-886-16	Sequence 16, Appli
35	71.5	7.1	1168	1	US-08-230-047-4	Sequence 4, Appli
36	71.5	7.1	1168	2	US-08-829-026A-5	Sequence 5, Appli
37	71.5	7.1	11464	4	US-08-991-840A-2	Sequence 2, Appli
38	71	7.1	927	1	US-08-507-431-5	Sequence 5, Appli
39	71	7.1	927	2	US-08-902-655A-5	Sequence 5, Appli
40	71	7.1	927	3	US-09-116-622-5	Sequence 5, Appli
41	71	7.1	927	4	US-09-219-277-5	Sequence 5, Appli
42	71	7.1	927	4	US-09-599-661-5	Sequence 5, Appli
43	71	7.1	2191	3	US-08-632-806A-6	Sequence 6, Appli
44	71	7.1	2192	1	US-08-273-538A-6	Sequence 6, Appli
45	70	7.0	9069	4	US-08-961-527-97	Sequence 97, Appli

ALIGNMENTS

RESULT 1
US-09-221-017B-726
Sequence 726, Application US/09221017B
Patent No. 6444799
GENERAL INFORMATION:
APPLICANT: Ross, Bruce C.
TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
NUMBER OF SEQUENCES: 1120
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221,017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998

No paper

:	ATTORNEY/AGENT INFORMATION:	
:	NAME: Monroy, Gladys H	
:	REGISTRATION NUMBER: 32,430	
:	REFERENCE/DOCKET NUMBER: 27340-20021..00	
:	TELECOMMUNICATION INFORMATION:	
:	TELEPHONE: 650-813-5600	
:	TELEFAX: 650-494-0792	
:	TELEX: 706141	
:	INFORMATION FOR SEQ ID NO: 726:	
:	SEQUENCE CHARACTERISTICS:	
:	LENGTH: 1974 base pairs	
:	TYPE: nucleic acid	
:	STRANDEDNESS: double	
:	TOPOLOGY: circular	
:	MOLECULE TYPE: DNA (genomic)	
:	HYPOTHETICAL: NO	
:	ANTI-SENSE: UNKNOWN	
:	ORGANISM: PORPHYROMONAS GINGIVALIS	
:	FEATURE:	
:	NAME/KEY: misc_feature	
:	LOCATION: 1...1974	
:	US-09-221-017B-726	
:		
:	Alignment Scores:	
:	Pred. No.: 1_97e-111 Length: 1974	
:	Score: 977.00 Matches: 191	
:	Percent Similarity: 98.96% Conservative: 0	
:	Best Local Similarity: 98.96% Mismatches: 0	
:	Query Match: 97.41% Indels: 2	
:	DB: Gaps: 0	
:		
:	US-10-008-355-2_COPY_522_712 (1-191) x US-09-221-017B-726 (1-1974)	
OY	1 SerLysserValIlealaaIaaArgAlaIleGlInAlaSPalaMetAlaSnalAatyr	20
Db	1156 TCCAAAGAGCGTAATAGCTGCTGCCGCATTACAGCCGATGGCATGCCATCCTAT	1215
OY	21 AAlallecluySglYkysArgleuphepheaIaglyLeuaRgUmetyrProglyarg	40
Db	1216 GCCATTAGAGAAGGCAAGCCCTCTTTCTTGGCGGTTCGTAAGATGAACCCGGAGCT	1275
OY	41 AlaleuproSeraspAlaasnPhetHrmetArgmetSerTyrglySerIleueySgltyr	60
Db	1276 GCTTGCGGAGCGATGGCAACTTCACCATGCGATGAGCTACGGCTCATCAAGGATAT	1335
OY	61 GlupeRodlnaspgIyalatPryfAsnTYrhIstHrThrglyVvalleunluyS	80
Db	1336 GAACCGAGAGCGGTGCTGCTGTCACATTCATCACAGCAGCAGGAGGCTATGGAGAG	1395
OY	81 GlinAsp-Pro-LysserAspGluPheAlaValcInGlnuSnlleuaspLeupheargt	100
Db	1396 CAGGATTCCTTAAGAGCGAGATGTTGCCGTACAGAGAAATPCTCGACCTTCCGCA	1455
OY	100 hrLysaenTYrglyArgTYrAlaGlnuSnGlyInleuhsIstleAlapheleuserrAsna	120
Db	1456 CCAAAAACTATGGTCGTATAAGCCGAAAGGGTCAGCTCATTCCTTCCATCGAACCA	1515
OY	120 snaspiIethrInglylYAsnseryGlySerProvalPheaspLysasnGlyArgleulleg	140
Db	1516 ACAGCAATCACGGGGGCTTAAGTCCCGGTAAGCCCGGTATTCATAGAAAGCCGCTGATCG	1575
OY	140 lylenuAlapheaspGlyAsnTrpGlnAlaMetserGlyspIleelugPhegluproasPL	160
Db	1576 GTCTTGCTTTGATAGGCAACTGGGAAGCATATAGTGTGATCATCGAGTTCCAGACCGATC	1635
OY	160 euglinatrgHrIleserValAspIleatrgTYrValleuphemeticIleaspLystrypGly	180
Db	1636 TGCAAGCCACAATACAGCTGGATCCGCTAGCTTCTTCATGATGTACAAATAGGCTC	1695
OY	180 lngysproArgleulleGlnGlnleuleysleutle	191
Db	1696 AGTGGCCCCGCTCATCCAAAGACTCAAAGTGAAT	1730

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```
Db 114 ATTGAAGGGCACAAGCAGCTACATTCAGCTCTTGGAAATGATGGCGATCAT 173
Oy 42 LeuproSerAspAlaAsnPhenTherMetSerTyrGlySerIleGlyTyrGlu 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174 CAATTCGGGATGCTTACTGACACTTCCTTTCACCTATGTCAGAGTGAGGGCTATTCA 233
Oy 62 ProGlnAspGlyAlaTrrPyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGln 81
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 234 CCCCGTAGAATGTTAGTACGACATCAACACCATGATGGTGTGATGAAAAAGAA 293
Oy 82 AspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThrLys 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 GATCCCGGATTAATGGGAAATTGTAGTCATCCCAAGCTCAAGCCGTAATACAGCGTAA 353
Oy 102 AsnTyrGlyArgTyrAlaGlu---AsnGlyGlnLeuHisIleAlaPheLeuSerAsn 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 354 GACTTCGGGCGCTTATGCCATGCGACGGTCCGATGCCCTGTGGCTTTGGCCACACCA 413
Oy 121 AspIleThrGlyLysSerGlySerProValPheAspLysAsnGlyArgLeuIleGly 140
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 414 CATACACCGGGCGCACTCAGCAGCTCCGTCATGAAATGCCAAGCGCACTGATCGGT 473
Oy 141 LeuAlaPheAspGlyAsnTrrPglAlaMetSerGlyAspIleGluPheGluProAspLeu 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 474 CTCACACTTCGATCTGACCTGAGGAGGAGCGGTGGCGCATCTCAGTATCTGGCCGACTAC 533
Oy 161 GlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTrpGlyGln 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 534 CAGCCGACGATATGTTGATATTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 593
Oy 181 CysProArgLeuIleGlnGluLeuLysLeuIle 191
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 594 TGCCAAAGCGCTGTGATGAATGAATATCCTT 626

RESULT 3
US-09-058-411-1
; Sequence 1, Application US/09058411
; Patent No. 6171854
; GENERAL INFORMATION:
; APPLICANT: GALLER, Ricardo
; TITLE OF INVENTION: VACCINES AGAINST INFECTIONS CAUSED BY YF
; TITLE OF INVENTION: VIRUS: YF INFECTIONS CDNA, METHOD FOR PRODUCING A
; TITLE OF INVENTION: RECOMBINANT YF VIRUS FROM THE YF INFECTIONS CDNA AND
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN
; ADDRESSEE: Intellectual Property Group of
; ADDRESSEE: PILLSBURY MADISON & SUTRO LLP
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058.411
; FILING DATE: April 10, 1998
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: BR 9701774.4
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: WHITE JR, PAUL E
; REGISTRATION NUMBER: 32011
; REFERENCE/DOCKET NUMBER: 31329/251760
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
```

```
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10862 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: YF1v5.2/DD
;
US-09-058-411-1
;
Alignment Scores:
Pred. No.: 1.53 Length: 10862
Score: 89.50 Matches: 39
Percent Similarity: 39.22% Conservative: 21
Best Local Similarity: 25.49% Mismatches: 46
Query Match: 8.92% Indels: 47
DB: Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x US-09-058-411-1 (1-10862)
Oy 19 AlaTyrAlaIleGlnLysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetYrPro 38
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4742 GCTTCCTGTCAGGAATGCAAGAGTGG----- 4771
Oy 39 GlyArgAlaLeuProSerAspAlaAsnPhenTherMet---SerTyrGlySerIle 57
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4772 -----ATTCATCTTGGGCTTCAGTAAAGAAAGACCTTGTCCTATATGCT 4816
Oy 58 LysGlyTyrGluProGlnAspLysAlaTrrPyrAsnTyrHis----- 71
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4817 ---GGCTCATGGAATGTGACAGCAGATGCGATGAGAGAGAGGTCCAGTTGATCGCG 4873
Oy 72 ---ThrThrGlyLysGlyValLeuGluLysGlnAspProLysSerAspLeuPheAlaVal 90
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4874 GCTGTCCAGGAAGAACCTGCTCAACCTCAGACAAACCGAGC----- 4918
Oy 91 GlnGluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGly 110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4919 -----TTGTTCCAAAGTGAGGATGGGGG-----GAAATCGGG 4951
Oy 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGlyLysAsnSerGlySerPro 130
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4952 GCTGTCGCTCT-----GACTATCCGAGTGGCACTTCAGGATCTCT 4993
Oy 131 ValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrrPglAlaMet 150
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4994 ATTGTTAACAGACGACGAGAGGTGATGGCTGTACGGCAATGCG-----ATCCTT 5044
Oy 151 SerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5045 GTGCGTGACAACTCTTCTGTCGCCCATATGCCAGACT 5083

RESULT 4
US-09-134-001C-1520
; Sequence 1520, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134.001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064.964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055.779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1520
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
```

US-09-134-001C-1520

Alignment Scores:

Pred. No.: 0.412 Length: 1083
Score: 83.00 Matches: 34
Percent Similarity: 35.25% Conservative: 15
Best Local Similarity: 24.46% Mismatches: 46
Query Match: 8.28% Indels: 44
DB: 4 Gaps: 6

US-10-008-355-2_COPY_522_712 (1-191) x US-09-134-001C-1520 (1-1083)

OY 57 ILeLysGlyLysArgLeuPhePheAlaThrPheAsnTyrHisThrGlyLysGly 76
DB 58 ATGAAGGGGCGCATCCATGAAACATCAAGAAACCTTGATTAAAGTAAAGTCA 117
OY 77 ValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAsp 96
DB 118 TTGCTAAAGAGTATGATCTTACAGGTGAGAAATTT-----GAAGCTTAATCGAT 168
OY 97 LeuPheArgThr-----LysAsnTyr-----GlyArgTyrAlaGluAsn 109
DB 169 TTGCTATGACATTTAAAAATATATAACACAGCACACCATCGATATTATTAAGGCT 228
OY 110 GlyGlnLeu----- 112
DB 229 AAGACATTTGCTTACTCTTCGAAAGACATCTACTCGAGCGGCGCATTTACAGTC 288
OY 113 -----HisIleAlaPheLeuSerAsnAspIleThrGlyGly 125
DB 289 GCATCTATTGATCTAGTGCACACCCCTGAATTTTGGGAAAAAAGATTCATTAAGTA 348
OY 126 AsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGly 145
DB 349 AAAAAAGATCT-----CTTGAGCATCTGCTAAGTTTGGCGAAGATGTTTATGCA 402
OY 146 AsnTyrGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164
DB 403 -----ATTGATTTAGAGGTTTTCACAAAAAAGCTGT 435

RESULT 5

US-09-660-587-3
; Sequence 3, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 3
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide
; OTHER INFORMATION: nucleic acid sequence of p28-5
US-09-660-587-3

Alignment Scores:
Pred. No.: 0.521 Length: 849
Score: 81.00 Matches: 43
Percent Similarity: 39.38% Conservative: 20
Best Local Similarity: 26.88% Mismatches: 67
Query Match: 8.08% Indels: 30
DB: 4 Gaps: 7

US-10-008-355-2_COPY_522_712 (1-191) x US-09-660-587-3 (1-849)

OY 23 GluLysGlyLysArgLeuPhePheAlaThrPheAsnTyrHisThrGlyLysGly 42
DB 172 GAGAAAAAAGACACACTGCTATGCTTAAAGAAAACTGGCGAGAGATGCAATA 231
OY 43 ProSer-----AspAlaAsnPheThrMetArg-----MetSerTyrGlySer 56
DB 232 TCTACTCAAGGTCAGATGATTAATTTTACATTCGAATTTCTCATCAAGATGCAAGC 291
OY 57 ILeLys-----GlyTyrGluProGlnAspGlyAlaThrPheAsnTyrHisThrGly 74
DB 292 AACCAAGTTTAAAGGTTGTCAGTACCTATTGCT-----TACTCGATTAGC 336
OY 75 LysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIle 94
DB 337 AGTCCAAAGATGAA-----GTTGACATGCTCT 363
OY 95 LeuAspLeuPheArgThrLysAsnTyrGlyArgTyrAlaGluAsnGlyLysIle 114
DB 364 TATGAAAGATTTGATGTAAGAAATCCAGTGATATTACAAAACGGCTTACAGTAT 423
OY 115 AlaPheLeuSerAsnAspIleThrGlyGlnAsnSerGlySerPro----- 130
DB 424 TGTGCTTATCTCATCAAGATGATCGGATGATGACATGACATGACAAATTT 483
OY 131 ValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGlyAsn-----TrpGlu 148
DB 484 GTATATTTAATTAAGAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTA 543
OY 149 AlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSerValAspIle 168
DB 544 ACAGCAACCAAAAAATATACCTCTCTCTTACATATGTCAGGATTTGTAATTTTA 603

RESULT 6

US-09-261-358A-3
; Sequence 3, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 3
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide
; OTHER INFORMATION: nucleic acid sequence of Eca28Ssa2
US-09-261-358A-3

Alignment Scores:
Pred. No.: 0.521 Length: 849
Score: 81.00 Matches: 43
Percent Similarity: 39.38% Conservative: 20
Best Local Similarity: 26.88% Mismatches: 67
Query Match: 8.08% Indels: 30
DB: 4 Gaps: 7

US-10-008-355-2_COPY_522_712 (1-191) x US-09-261-358A-3 (1-849)

OY 23 GluLysGlyLysArgLeuPhePheAlaThrPheAsnTyrHisThrGlyLysGly 42
DB 172 GAGAAAAAAGACACACTGCTATGCTTAAAGAAAACTGGCGAGAGATGCAATA 231


```

Db 205702 -----CGGGTGGGCGACGACGTATCGACCTG 2057720
Oy 98 PheargthrLysasnYrGlyArgTYrAlaGlnasnGlycInLeuHisIleAlaPheLeu 117
    |||
    |||:::||||
Db 205729 GTGTCCACCGGTGCTCCGGTAATATCTTCCTCCCGACAG----- 2057700
Oy 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySer-----ProValPheAspLys 134
    |||
    |||:::||||
Db 205771 -----ACCATCCACCAAGGACCGCTTCCAGTACATCGGGCGCGCTGACAAAT 205821
Oy 135 AsnGlyArg-----LeuIleLeuAlaPheAspGlyAsn 146
    ::|
    ::|::|
Db 205822 TCCAATCGCGGGGTGGCCGCATTCGCCACGAGAAAGATCGGCTTGCTGCTGCACGACGACC 205881
Oy 147 TrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIle 164
    ::|
    ::|::|
Db 205882 GCGCAAGCGGTGGGG-----CTGGACCCCGCTTGCAACCGTTGGTGC 205926
RESULT 9
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007 / 00
; CURRENT APPLICATION NUMBER: US/09/103, 840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
  Align. No.: 1.63e+05 Length: 4411529
  Score: 77.50 Matches: 40
  Percent Similarity: 32.91% Conservative: 12
  Best Local Similarity: 25.32% Mismatches: 45
  Query Match: 7.73% Indels: 61
  DB: 4 Gaps: 8
US-10-008-355-2_COPY_522_712 (1-191) x US-09-103-840A-1 (1-4411529)
Oy 36 MetYrProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArg---MetSerTYr 54
    ::|
    ::|::|
Db 205382 CTGTATCGC-----ACGGCAATGTGACTACCGCGGTATCACCATT 205423
Oy 55 GlySerIleLysGlyTYrGluProGlnAsp----- 64
    |||
    |||::|
Db 205424 GCGAAGGTTACTGCGCGTGCACCCACGACGAGGCGCAGCAGTACGATGATGATCGCC 205483
Oy 65 -----GlyAlaTrpTYrAsnTYrHisThrThrGlyLysGlyVal 77
    |||
    |||::|
Db 205484 AGCAACTACAAATCCCGCTGCATGCTCGGCGCAACCTGATTCGCTGCA----- 205534
Oy 78 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGlnAsnIleAsnAspLeu 97
    |||
    |||::|
Db 205535 -----GCGGTGGGCGAGCAGTACATCGACCTG 205561
Oy 98 PheargThrLysAsnTYrGlyArgTYrAlaGlnasnGlycInLeuHisIleAlaPheLeu 117
    |||
    |||:::||||
Db 205562 GTGTCCACCGGTGCTCCGGTAATATCTTCCTCCCGACAG----- 205603
Oy 118 SerAsnAsnAspIleThrGlyGlyAsnSerGlySer-----ProValPheAspLys 134
    |||
    |||::|

```

```

Db 205604 -----ACCATCACCAAGGCGACCGCTTCCAGTCAGATCGGGCGCGGCTGCGACAAAT 205654
Oy 135 AsnglyArg-----LeuileglyLeualaIheaspGlyAsn 146
::: ||| ||||||| |||
Db 205655 TCCAAATCGCGGCTTGCGCCGCAATTCGCCACGAGAGATCGGCTTGCTGCTGCAGAGACC 205714
Oy 147 TrpGluAlaMetserGlyAspIleGluIupheGluIuproAspIeuGlnIupArgThrIle 164
:::|||||::: ||| ||| |||||||:::
Db 205715 CGCGCAAGCGGTGGGTGG-----CTGGAGCCGCGGCTTGCAACGCTTGTC 205759

RESULT 10
US-09-441-340-2
; Sequence 2, Application US/09441340
; Patent No. 6448476
; GENERAL INFORMATION:
; APPLICANT: Barry, Gerard F.
; TITLE OF INVENTION: Phosphonate Metabolizing Plants
; FILE REFERENCE: 38-21(15303)
; CURRENT APPLICATION NUMBER: US/09/441,340
; CURRENT FILING DATE: 1999-11-16
; EARLIER APPLICATION NUMBER: 60/108,763
; EARLIER FILING DATE: 1998-11-17
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 11672
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-441-340-2

Alignment Scores:
Pred. No.: 106 Length: 11672
Score: 75.00 Matches: 28
Percent Similarity: 40.43% Conservative: 10
Best Local Similarity: 29.79% Mismatches: 42
Query Match: 7.48% Indels: 14
DB: 4 Gaps: 3

US-10-008-355-2.COPY_522_712 (1-191) x US-09-441-340-2 (1-11672)
Oy 5 IleAlaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaSnAlaIArgAlaIleGlu--- 23
||| ||| |||||||::: |||
Db 1034 ATCACAGCAAAACCTGAACCCGCAATGGACGCCGCTTCTCAGATATGGAGAAGAAAGCT 1093
Oy 24 -----LysGlyLysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAla 41
:::|||||::: ||| ||| ||| ||| ||| |||
Db 1094 GGGCGTGAAGGTCAACGCGCTTCTTGCGCCGCACTACGGCGCATTTATCCAGGGGAGATCG 1153
Oy 42 LeuProSerAspAlaSnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGlu 61
||| ||| ||| ||| |||
Db 1154 CTTCAATTAAGTGTGATATCGCTCGTACGGCAATCTGCGCATGTGAACGGTGTGATCG 1213
Oy 62 ProGluAspGlyAlaThrTyrAsnTyrHisThrThrglyLysGly----- 76
||| ||| ||| ||| ||| ||| ||| |||
Db 1214 CGCGCA-----TGGCCAGGCTCTTCGCCCAAGACGCTGCGGCGGATGATCGCCGG 1264
Oy 77 -----ValLeuGluLysGlnAspProLysSerAspGlu 87
:::|||||::: ||| ||| ||| ||| |||
Db 1265 TTACTGTGAGCGTGTGATCGTCAACAAGACACTCCATCA 1306

RESULT 11
US-09-454-721A-3
; Sequence 3, Application US/09454721A
; Patent No. 6296854
; GENERAL INFORMATION:
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Bruce J. Crise
; TITLE OF INVENTION: Live Attenuated Venezuelan Equine Encephalitis Vaccine
; FILE REFERENCE: Army 146
; CURRENT APPLICATION NUMBER: US/09/454,721A
; CURRENT FILING DATE: 1999-12-07

```


PRIOR APPLICATION NUMBER: US 60/111,330
PRIOR FILING DATE: 1998-12-07
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Microsoft Word
SEQ ID NO 3
LENGTH: 11459
TYPE: DNA
ORGANISM: Venezuelan Equine Encephalitis Virus
FEATURE:
US-09-454-721A-3

Alignment Scores:
Pred. No.: 119 Length: 11459
Score: 74.50 Matches: 40
Percent Similarity: 37.42% Conservative: 21
Best Local Similarity: 24.54% Mismatches: 57
Query Match: 7.43% Indels: 45
Gaps: 11

US-10-008-355-2_COPY_522_712 (1-191) x US-09-454-721A-3 (1-11459)

QY 38 ProGlyArg-----AlaLeuProSerAspAlaAsnPhetArg 51
Db 7895 CCAGGCACAGACACCGCATGTCATGAAATGCAATCTGCACAGACGTTCCCAATC--- 7951

QY 52 MetSerTyGlySerIleTyGlyTyGlyProGlnAspGlyAla-----TrpTyrAsn 69
Db 7952 ATGTTGGAGGGAAGATTAACGGCTACGCTTGCTGCTGGAGGCAAGTTATTCAGCCG 8011

QY 70 TyrHisThrThyGlyLys-----GlyValLeuGlnLysGlnAsp 82
Db 8012 ATGCATGTGAGGAGGCAAGATTCACACAGACGTTTCGGCCGCTTAAGCAAGAAGCA 8071

QY 83 ProLysSerAsp---GluPheAla---ValGlnGluAsnIle---LeuAspLeuPheArg 99
Db 8072 TCCAAATACGATCTGAGTATGCACAGATGTGCCACAGAACATCGCGCCGATCATTCAAA 8131

QY 100 -----ThyLysAsnTyGlyArgTyGlyAla-----Glu 108
Db 8132 TACACCCATGACAAACCCCAAGGCTATTACAGCTGGCATCATGACGACGATCCAAATATGAA 8191

QY 109 AsnGlyGlnLeuHisIleAlaPheLeuSerAsnAspIleThyGlyLysAsnSergly 128
Db 8192 AATGGGCGGTTTCACGGTg-----CCGAAAGAGTTGGGGCCCAAGGACACAGCGGA 8242

QY 123 SerProValPheAspLysAsnGlyArgIleGlyLeuAlaPheAspGly----- 145
Db 8243 CGACCCATTCTGATTAACGAGGAGCGGTGCTGCTATTGCTGGGAGGTGTGATGAA 8302

QY 146 -----AsnTrpGluAlaMetSerglyAspIleGluPhe 156
Db 8303 GGATCTAGACAGCCCTTTCATGCTCATGTGGAACGAGAGGAGATTACCGTGAAGTAT 8362

QY 157 GluProAsp 159
Db 8363 ACTCCGGAG 8371

RESULT 12
US-08-433-522A-9
Sequence 9, Application US/08433522A
Patent No. 6013514
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: THOMAS, Wayne
APPLICANT: YANG, Yan Ping
APPLICANT: LOOSMORE, Sheena
APPLICANT: SIA, Dwo Yuan Charles
APPLICANT: KLEIN, Michel
TITLE OF INVENTION: HAEMOPHILUS OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESS: Sim & McBurney
STREET: 6TH Floor, 330 University Avenue

CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,522A
FILING DATE: 12-SEP-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-434 MIS:jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2989 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 390..2768
US-08-433-522A-9

Alignment Scores:
Pred. No.: 21.6 Length: 2989
Score: 74.00 Matches: 39
Percent Similarity: 33.88% Conservative: 23
Best Local Similarity: 21.31% Mismatches: 55
Query Match: 7.38% Indels: 66
Gaps: 9

US-10-008-355-2_COPY_522_712 (1-191) x US-08-433-522A-9 (1-2989)

QY 16 MetAlaAsnAlaTyGlyAlaIleGlyLysArgLeuPhePheAlaIleuArgGlu 35
Db 1383 TTAGCATATACCTTTGTTGTTGATGCTGGACGACGTTTAACGTTCGCCAAGCTTGC--- 1439

QY 36 MetTyProGlyArgAlaLeuProSerAspAlaAsnPhetMetArgMetSergly 55
Db 1440 ---TTGAGGAATAATACGCTTCTGCTGATGACTTTACGACGAATG----- 1487

QY 56 SerIleLysGlyTyGlyGluProGlnAspGlyAlaTrpTyrAsnTyHisThrThyGlyLys 75
Db 1488 -----CGACACACAGAGAGAACTTGATATAATTCACAAATTAAGTTA 1532

QY 76 GlyValLeuGlnLysGlnAspProLysSerAspGluPheAlaValGlnGlnAsnIleLeu 95
Db 1533 GGAATAATTCGCTTAGAT-----CGTACAGCTTCTTGAAACAGCTTAAACCGAAT 1586

QY 96 Asp-----LeuPheArgThrLys----- 101
Db 1587 GATCCATCATGATGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1646

QY 102 -----AsnTyGlyArgTyGlyAlaGlnAsnGly----- 110
Db 1647 ACGGATGATCACTTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1700

QY 111 -----GlnLeuHisIleAlaPhe 116
Db 1701 CAACAAGATTAATTAACAGATATTTCTTGGCAACAGCGCGCGCATTAAGTATGCTGCT 1760

QY 117 LeuSerAsnAsnAspIleThyGlyGlyAsnSergly-----SerProValPheAspLys 134
Db 1761 ACGAAATAATGATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1820


```

REGISTRATION NUMBER: 24_973
REFERENCE/DOCKET NUMBER: 1038-732 MIS-Jb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2989 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 390..2768
US-08-942-046-9

Alignment Scores:
Pred. No.: 21.6 Length: 2989
Score: 74.00 Matches: 39
Percent Similarity: 33.88% Conservative: 23
Best Local Similarity: 21.31% Mismatches: 55
Query Match: 7.38% Indels: 66
DB: Gaps: 9

US-10-008-355-2_COPY_522_712 (1-191) x US-08-942-046-9 (1-2989)
QY 16 MetAlaAsnAlaTyrAlaIleGluLysGlyLysAlaGluPhePhaIleGlyLeuAtrGlu 35
Db 1383 TTACGCATTAACCTTTGTGGTGATGGTCAGCAGCACTTTACTGTTGCCAATTCCG--- 1439
QY 36 MetArgProGlyArgAlaLeuProSerAspAlaAsnPhetMetArgMetSertArgLy 55
Db 1440 ---TTTAGAGAAGAAATACCCTTTCTCGTCATAGACTTTACGCAGGAATG----- 1487
QY 56 SerIleuSeqLyTyrgLuPProGlnAspGlyAlaTrpTyraSnTyHISThrThelyLys 75
Db 1488 -----CGACAACAAGAGAACCTTGGTATAATTCACAATTAGTTAGTACTTA 1532
QY 76 GlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleu 95
Db 1533 GGAAAATTTGGCTTAGT-----CGTACAGCTTTCTTGGAACAGTTGAAAAACCGAATT 1586
QY 96 Asp-----LeuPheArgThrLys----- 101
Db 1587 GATCCATCATCATGTAGCAGATGATGAAGTGAATGTCGTTATTAAGTCAAGAACGTAAAC 1646
QY 102 -----AsnTyrgLyaaTygTyAlaGluAsnGly----- 110
Db 1647 ACGGGTACTATCAACTTTTGGTATGGTTACGGT-----ACAGAAAGTGTAATTCAGTTAT 1700
QY 111 -----GlnLeuHisIleAlaPhe 116
Db 1701 CAACAGATATTAAACAAGATATTCTTGGGAACAGGGGGCAGTAGATACATACCTTGT 1760
QY 117 LeuSerAsnAsnAspIleThrGlyGlyAsnSergLy-----SerProValPheAspLys 134
Db 1761 ACGAAAATATATTATGAGTACGAGTCTCAATTGGGTATATACCGAACCCCTAATTTCTAAA 1820
QY 135 AsnGlyArgLeuIleGlyLeuAlaIleAspGlyAsnTrpGluAlaMetSergLyAspIle 154
Db 1821 GATGGTCTAAGCTCTGCT-----GGAATATAT 1847

```

```

: APPLICANT: Yang, Yan-ping
: APPLICANT: Klein, Michel H.
: TITLE OF INVENTION: Protective Recombinant Haemophilus Influenzae High
: TITLE OF INVENTION: Molecular Weight Proteins
: FILE REFERENCE: 1038-861 M15:b
: CURRENT APPLICATION NUMBER: US/09/206,942
: CURRENT FILING DATE: 1998-12-08
: EARLIER APPLICATION NUMBER: 09/167,568
: EARLIER FILING DATE: 1998-10-07
: NUMBER OF SEQ ID NOS: 95
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 52
: LENGTH: 2934
: TYPE: DNA
: ORGANISM: Haemophilus influenzae
US-09-206-942-52

```

Alignment Scores:	
Pred. No.:	24.3
Score:	73.50
Percent Similarity:	37.148
Best Local Similarity:	26.578
Query Match:	7.338
DB:	4
Gaps:	6
Length:	2933
Matches:	40
Conservative:	12
Mismatches:	69
Indels:	19

US-10-008-355-2_COPY_522_712 (1-191) x US-09-206-942-52 (1-2934)

```

Oy 42 leProSeArPaLaSnPrHeImeArGmet-----SerGlySerIleLeuS 58
    ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 631 TTAAGCAtGTGTGGAGAAATTTACCTTTATTAAATATGCGAGCAAGCATTCAGCAAC 690
Oy 59 GluTyGlnProGlnAsrGluAlaTrpTyGAsnTyrGlnStpThrStcLysGluValLeu 78
    ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 691 TCCAGTATTCCAGCAATTTTGGCGAGAGTAATTCAC-----GCGAAGAT----- 738
Oy 79 GluLysGlnAsrProLysSerAspGluPheAlaValcGlnLysnIleLeuAspLeuPhe 98
    ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 739 -----AATGAATGAATTAATTAATTTGGTAATATCCCAAGCGTAAATTT 783
Oy 99 ArgThrLysAsnTyGluGluGluTyGAlaGlnLysGlnLeuIleStpLeuIleAspLeuSer 118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 784 AGGTTAAACCAAAATGAGAGACACTCCCTAGACAGCACTCCCAATTTAGATTTATCT 843
Oy 119 AsnAsnAspIleThrGlnGlyAsnSerGlySerProValPheAspLysAsnGluArgLeu 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 844 AATATTTCGGCTACACTGGCGGA-----GGTTCGTGTTTTCGATATATACCTTAACCTT 897
Oy 139 IleGlyLeuAlaPheAspGlyAsnTrpGluAlaIleMetSerGlyAspIleGluIlePro 158
    ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 898 TGGGGT-----AAAGGAGCTAGACTAAAGATGATTCATTAACCTTTCTAGAC 945
Oy 159 AspLeuGlnArgThrIleSerValAspIleArg---TyrValLeuPheMetIleAspLys 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 946 GGCCTCATCTTAACTTAATTCCTCATGTTCCGAGACAGTAATATGCTTTTAAATCAATAA 1005

```

Search completed: December 21, 2002, 02:28:25
Job time : 2265 secs

RESULT 15
US-09-206-942-52
; Sequence 52, Application US\09206942
; Patent No. 6432659
; GENERAL INFORMATION:
; APPLICANT: LOOMISMORE, Sheena M.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 21/ 2002, 01:47:03 ; Search time 2159 Seconds

(without alignments)
1432.764 Million cell updates/sec

Title: US-10-008-355-2_COPY_522_712

Perfect score: 1003

Sequence: 1 SKSVTAARAIQADAMANAY.....LFMIDKQCPRLIQELKLI 191

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=framer.p2n.model -DEV=xlh
-O=/con2_1/USPRO.spool/US10000355/runat_17122002_112705_16942/app.query.fasta_1_327
-DB=EST-QFMT-fastlap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human4.cdi -LIST=45
-DOCAALIGN=200 -THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10000355.ecgn.1.1.763@runat_17122002_112705_16942 -NCPU=6 -TCPU=3
-NO_XIPXY -NO_MMAP -LANG=ENGLISH -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estinu: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_estcl: *
10: gb_estl2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_esttom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pin: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_trod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
c 1	314	31.3	426	17	BH400391	BH400391 AG-ND-147
2	136	13.6	822	17	BH371846	BH371846 AG-ND-162
3	90.5	9.0	1102	13	BM562210	BM562210 AGENCOURT
4	89.5	8.9	630	12	BC491629	BC491629 602535962
5	89.5	8.9	667	13	BI391728	BI391728 pppin.pk0
6	89.5	8.9	701	10	BE268652	BE268652 601125126
7	89.5	8.9	710	14	BM638106	BM638106 K-EST0114
8	89.5	8.9	852	12	BQ215538	BQ215538 AGENCOURT
9	89.5	8.9	853	12	BE799506	BE799506 601589315
10	89.5	8.9	881	13	BM456284	BM456284 AGENCOURT
11	89.5	8.9	899	13	BM449938	BM449938 AGENCOURT
12	89.5	8.9	904	14	BO645307	BO645307 AGENCOURT
13	89.5	8.9	942	14	BO652808	BO652808 AGENCOURT
14	89.5	8.9	943	12	BG174574	BG174574 602334417
15	89.5	8.9	960	14	BO646879	BO646879 AGENCOURT
16	89.5	8.9	982	14	BO645695	BO645695 AGENCOURT
17	89.5	8.9	1004	14	BO606048	BO606048 AGENCOURT
18	89.5	8.9	1015	14	BO335052	BO335052 AGENCOURT
19	89.5	8.9	1027	14	BM928590	BM928590 AGENCOURT
20	89.5	8.9	1031	13	BM550804	BM550804 AGENCOURT
21	89.5	8.9	1036	14	BO957384	BO957384 AGENCOURT
22	89.5	8.9	1039	14	BM911788	BM911788 AGENCOURT
23	89.5	8.9	1045	14	BM914762	BM914762 AGENCOURT
24	89.5	8.9	1052	14	BM809287	BM809287 AGENCOURT
25	89.5	8.9	1074	14	BM909854	BM909854 AGENCOURT
26	89.5	8.9	1092	14	BM908687	BM908687 AGENCOURT
27	89.5	8.9	1097	13	BM455151	BM455151 AGENCOURT
28	89.5	8.9	1022	11	BC014232	BC014232 Homo sapi
29	88.5	8.8	656	13	BO918792	BO918792 602819275
30	86.5	8.6	947	14	BO649391	BO649391 AGENCOURT
31	85.5	8.5	759	10	BE335285	BE335285 601058826
32	85.5	8.5	1062	13	BM464181	BM464181 AGENCOURT
33	84.5	8.4	428	13	BI080699	BI080699 602878440
34	83.5	8.3	412	12	BF841384	BF841384 RC2-HT107
35	83.5	8.3	694	9	AU131725	AU131725
36	83.5	8.3	743	9	AU134050	AU134050
37	83.5	8.3	861	13	BI116059	BI116059 602866455
38	83.5	8.3	960	14	BO892534	BO892534 AGENCOURT
c 39	83	8.3	571	12	BG602643	BG602643 EST01733
40	83	8.3	655	13	BI334196	BI334196 602997681
41	83	8.3	1385	14	BM911372	BM911372 AGENCOURT
42	82.5	8.2	412	10	BB674829	BB674829
43	82.5	8.2	491	9	AL034929	AL034929 m8708a54
44	82.5	8.2	558	13	BM204311	BM204311 C0267F10-
45	82	8.2	893	12	BF699325	BF699325 602125888

ALIGNMENTS

RESULT 1
BH400391/c
LOCUS
DEFINITION AG-ND-147H4.TF ND-TAM Anopheles gambiae genomic clone AG-ND-147H4,
DNA sequence.
ACCESSION BH400391
VERSION BH400391.1 GI:17346607
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
Anopheles.
REFERENCE 1 (bases 1 to 426)
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

KEYWORDS	EST.
SOURCE	
ORGANISM	
REFERENCE	human
AUTHORS	Homo sapiens
TITLE	Eukarya: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.
JOURNAL	1 (bases 1 to 1102)
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D.

FEATURES	SOURCE	Location/Qualifiers
		1. 1162
		/organism="Homo sapiens"
		/db_xref="taxon:3606"
		/clone IMAGE:5480814"
		/clone_id="NH_MGC_41"
		/tissue_type="amelanotic melanoma, cell line"
		/lab_host="DH10B (phage-resistant)"
		/note="Jorgan: skin; Vector: pORB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGAGGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA Synthesis Kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NH_MGC Library."
BASE COUNT		301 a 272 c 304 g 215 t 10 others
ORIGIN		

Oy 125 ---GIYASnSerGISePrOvalPheSpIySaSnGIyAGLeuILeGIYleuNlaPhe 143
 ||||| ||| ||||| ||||| : : : ||||| |||||
 Db 410 TGTGGAAATGACGTGGAACTGCTTTTACCAAGAAVGGAAATGGATGGCATTTGCTTTC 469
 Oy 144 AspGIyASnTrpGluAlaMeSerGIyASpILeGIuPheGluProASpLeuInArThr 163
 ||||| : : ||| : : : ||| : : :
 Db 470 CGAATCCAGAAAGAACCTTTGGGGGGTACGGCCCTCTATCTCATGTCCCTGGTGAAGAT 529
 Oy 164 ILeSerValASpILeArGTyr 170
 : : : ||||| : : :
 Db 530 TGGCGATGGCACTGCACCTTC 550

RESULT 4	630 bp	mRNA	linear	EST 27-MAR-2001		
LOCUS	60253396221	NIH_MGC_41	Homo sapiens	CDNA clone IMAGE:4684958 5',		
DEFINITION	mRNA sequence.					
ACCESSION	BC491629					
VERSION	BC491629.1	GI:13453141				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 630)	NIH-MGC	http://mgc.ncl.nih.gov/	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999)
	Contact: Robert Strausberg, Ph.D.	Email: cgapbs-remail.nih.gov	Tissue Procurement: DCTD/DTP	
	CDNA Library Preparation: Ling Hong/Rubin Laboratory	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)	DNA Sequencing by: Incyte Genomics, Inc.	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.llnl.gov
	Plate: L10CM1493	row: 1	column: 15	High quality sequence stop: 630.
FEATURES	Location/Qualifiers			
source				1..630

```

Db 207 GTGAAGACACCTTCCTCCAGAGCCTGACCCCAAGCTGCTCATCGCTGTCCTG 266
      :::: |||||||
      ::::|:::
Qy 53 ---SerTyrGlySerIleLysGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHis 71
      ||| |::: ||| ||||| ||| |||
Db 267 GACTCCTGCAGCACCAGCTAGCGAAGAGCCTTCTCC-----TATGATCGATGA 317
      ||| |::: ||| ||||| ||| |||
Qy 72 TTTTGTGlyLysGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln 91
      |||||||
Db 318 GGCACCTGGGAG-----AACTCCACCAT----- 341
      ||| |::: ||| ||||| ||| |||
Qy 92 GluAsnIleLeuAspLeuPheArgTrpTyrAsnTyrGly---ArgTyrAlaGluAsnGly 110
      ||| |::: ||| ||||| ||| |||
Db 342 -----AGCCGGTTTGAAACTAGCGAGAGCAAGTGTGCAGAG----- 377
      ||| |::: ||| ||||| ||| |||
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGly----- 124
      ||| |::: ||| ||||| ||| |||
Db 378 -----AACGATGTCATTTGGCTGCTTGGCATTTTGA 410
      ||| |::: ||| ||||| ||| |||
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143
      ||| |::: ||| ||||| ||| |||
Db 411 TGTGGAATGAGCTGGAATGCTCTTTACCAAGAAAGTGAAGTGGCATGCTTTC 470
      ||| |::: ||| ||||| ||| |||
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspGluGlnArgThr 163
      ||| |::: ||| ||||| ||| |||
Db 471 CGAATCCAGAAAGAGCCTTGGGGGCTAGGCCCTCTATCTCATGCTGCTGTAAGAAT 530
      ||| |::: ||| ||||| ||| |||
Qy 164 IleSerValAspIleArgTyr 170
      ||| |::: ||| ||||| ||| |||
Db 531 TCGCGAGTGGAGTCAACTTC 551
      ||| |::: ||| ||||| ||| |||

RESULT 5
BI391728 667 bp mRNA linear EST 06-AUG-2001
LOCUS
DEFINITION pgpin.pk005.f8 Normalized Chicken Pituitary/Hypothalamus/Pineal
      Library gailus.gailus cdna clone pgpin.pk005.f8 5' similar to
      gblAAH04057.1|AAH04057 (BC004057) Similar to transforming, acidic
      coiled-coil containing protein 2 [Mus musculus], mRNA sequence.
ACCESSION BI391728.1 GI:15085010
VERSION
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
      Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 667)
      Porter,T.E. and Cogburn,L.A.
      ESTs from Normalized Chicken Pituitary/Hypothalamus/Pineal cDNA
      library USDA/IFARS Animal Genome Project
      Unpublished (2001)
JOURNAL Contact: Larry A. Cogburn
COMMENT University of Delaware
      Townsend Hall, Newark, DE 19717, USA
      Tel: 302-831-1335
      Fax: 302-831-2822
      Email: cogburnudel.edu, www.chickest.udel.edu.

FEATURES
      source
      1..667
      /organism="Gallus gallus"
      /strain="Commercial broiler chicken"
      /db_xref="taxon:9031"
      /clone="pgpin.pk005.f8"
      /clone_lib="Normalized Chicken
      Pituitary/Hypothalamus/Pineal Library"
      /sex="Male and female"
      /tissue_type="Pituitary Gland/Hypothalamus/Pineal Gland"
      /dev_stage="Embryonic (dl2,dl4,dl9); post-hatch (w1,w3,w5
      ,w7,w9)"
      /lab_host="E. coli EMDH10B"
      /note="Vector: pCMVSPORT6; Library made from equivalent
      pools of total RNA isolated from each tissue at different
      ages. Single pass sequencing from 5'-end"

```

```

BASE COUNT 223 a 116 c 178 g 132 t 18 others
ORIGIN
Alignment Scores:
Pred. No.: 0.247 Length: 667
Score: 89.50 Matches: 28
Percent Similarity: 47.86% Conservative: 28
Best Local Similarity: 23.93% Mismatches: 54
Query Match: 8.92% Indels: 7
DB: 13 Gaps: 2

US-10-008-355-2_COPY_522_712 (1-191) x BI391728 (1-667)
Qy 6 ALaAlaAlaArgAlaIleGlnAlaAspAlaMetAlaAsnAlaTyrAlaIleGluLysGly 25
      ||||| ||| ::::|::: ||| |::: ||| |::: |||
Db 176 GCTGCACCTACGACTTCCCGACAGACAGATTTGGCCCAAGAAAGAAAGTGTCTGAGTGC 235
      ||| |::: ||| ||||| ||| |||
Qy 26 LysArgLeuPhePheAlaGlyLeuArgGluMetTyrProGlyArgAlaLeuProSerAsp 45
      ||| |::: ||| ||||| ||| |||
Db 236 AAAGAGAAATATGAGAAAGCAAGAGCAAGTGAATGAGAAATAGTTCACAA 295
      ||| |::: ||| ||||| ||| |||
Qy 46 AlaAsnPheThrMetArgMetSerTyrGlySerIleLysGlyTyrGluProGlnAspGly 65
      ||| |::: ||| ||||| ||| |||
Db 296 TATGAGAACAGCATTCCTCAGATGATAGC-----AAGCCGAGATGACAGACAG 349
      ||| |::: ||| ||||| ||| |||
Qy 66 AlaTrpTyrAsnTyrHisThrThrGlyLysGlyValLeuGluLysGlnAspProLysSer 85
      ||| |::: ||| ||||| ||| |||
Db 350 AAGTCTGTCTCCCATCAGCTGTTCAGCAGCTGATCGTGAAGAGAGCAGCGCTGGCG 409
      ||| |::: ||| ||||| ||| |||
Qy 86 AspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgTrpTyrAsnTyrGlyArg 105
      ||| |::: ||| ||||| ||| |||
Db 410 GATCTGAAGTGGAGAAATCTGCGCATCTTTTCAGC-----AGA 454
      ||| |::: ||| ||||| ||| |||
Qy 106 TyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsnAspIle 122
      ||| |::: ||| ||||| ||| |||
Db 455 TATGAAATAATGAGAGAGTGTGGAGGGGTTTCGAGAGACGAGAGAGTA 505
      ||| |::: ||| ||||| ||| |||

RESULT 6
BE268652 701 bp mRNA linear EST 13-JUL-2000
LOCUS
DEFINITION 601125126r1 NIH_MGC_8 Homo sapiens cdna clone IMAGE:344958 5',
      mRNA sequence.
ACCESSION BE268652
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 701)
      NIH-MGC http://mgs.nci.nih.gov/.
      National Institutes of Health, Mammalian Gene Collection (MGC)
      Unpublished (1999)
JOURNAL Contact: Robert Strausberg, Ph.D.
COMMENT Email: cga@bs-remail.nih.gov
      Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
      cDNA Library Preparation: Ling Hong/Rubin Laboratory
      cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
      DNA Sequencing by: Incyte Genomics, Inc.
      Clone distribution: MGC clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
      Plate: LILCM132 row: c column: 07
      High quality sequence start: 8
      High quality sequence stop: 678.

FEATURES
      source
      1..701
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="IMAGE:344958"
      /clone_lib="NIH_MGC_8"
      /tissue_type="Burkitt lymphoma"
      /lab_host="DH10B (phage-resistant)"
      /note="Organ: Lymph; Vector: pOTB7; Site_1: XhoI; Site_2:

```


ECORI; cDNA made by oligo-dt priming. Directionally cloned into EORI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 161 a 170 c 210 g 160 t
ORIGIN

US-10-008-355-2_COPY_522_712 (1-191) x BE268652 (1-701)

Alignment Scores:
Pred. No.: 0.269 Length: 701
Score: 89.50 Matches: 42
Percent Similarity: 38.32% Conservative: 22
Best Local Similarity: 25.15% Mismatches: 56
Query Match: 8.92% Indels: 47
DB: 10 Gaps: 8

19 AlaTyrAlaIleGluysGlylsArgluPhepHeAlaGlyLeuArsgLumetrypro 38
134 ACCTTGGGGGTAGAGAGGGCCGTATGCTTCGAGATGACATGCAATGAGAAATCTCC 193
39 GATGAlaLeuProSerAspAla-----AsnPhetHmetArgMet----- 52
194 GTGAGAACCTTCGCTCAACAGAGCCCTGACCCCACTGCTCCGTATCGCTGTCCTC 253
53 ---SerTyrGlySerIlelysglyTyrgluProGlnAspGlyAlaTrrpYrsAntyrHis 71
254 GACTCCTCGAGCAGCAGCAGCTAGCGAGAGCCTTCTCC-----TATGGCTATGGA 304
72 ThrThrglyLysGlyValleuGluysGlnAspProlyssSerAspGluPheAlaValGln 91
305 GGCACCTGGAG-----AGCCGTTTGAACCTACGGAGACAGATTTCACAG----- 328
92 GlnAsnIleLeuAspLeuPheArgrThrylsAsnTyrgly---ArgTyrAlaGluAsnGly 110
329 -----AGCCGTTTGAACCTACGGAGACAGATTTCACAG----- 364
111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrgly----- 124
365 -----AACGATGTCATTCGCTGCTTCGCGATTTCGAA 397
125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgrLeuIleGlyLeuAlaPhe 143
398 TGTGGAAATGACGTGACCTGCTTTACCAAGAAATGGAAGTGGGCAATGCTTTC 457
144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgrThr 163
458 CGAATCCAGAAAGAGAGCCCTGGGGGTACGGCCCTATCCATCATGTCTCTGGTGAAGAT 517
164 IleSerValAspIleArgrTyr 170
518 TGGCAGTGGAGCTTCACCTTC 538

RESULT 7
LOCUS BM838106 710 bp mRNA linear EST 06-MAR-2002
DEFINITION K-EST0114293 S9SNU601 Homo sapiens cDNA clone S9SNU601-73-C05 5',
mRNA sequence.
ACCESSION BM838106
VERSION BM838106.1 GI:19194515
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 710)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.U., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS

Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 73 row: C column: 05
High quality sequence stop: 710.

FEATURES
source

Location/Qualifiers
1..710

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S9SNU601-73-C05"
/clone_11b="S9SNU601"
/sex="M"
/tissue="type="Ascites"
/cell_line="Epithelial"
/cell_line="SNF-601"
/lab_host="Top10F"

/note="Organ: Stomach; Vector: pME18-FL3; Site_1: XhoI; Site_2: XhoI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including SfiI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized with Superscript II using SfiI oligo-dt primer. After first strand synthesis, RNA was degraded by NaOH treatment and cDNA was amplified by PCR reaction. The PCR products were digested with SfiI and cloned into DraIII- digested pME18-FL3 vector. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 163 a 182 c 201 g 164 t
ORIGIN

Alignment Scores:
Pred. No.: 0.275 Length: 710
Score: 89.50 Matches: 42
Percent Similarity: 38.32% Conservative: 22
Best Local Similarity: 25.15% Mismatches: 56
Query Match: 8.92% Indels: 47
DB: 14 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x BM838106 (1-710)

19 AlaTyrAlaIleGluysGlylsArgluPhepHeAlaGlyLeuArsgLumetrypro 38
115 ACCTATGGGGGTAGAGAGGGCCGTATGCTTCGAGATGACATGCAATGAGAAATCTCC 174
39 GATGAlaLeuProSerAspAla-----AsnPhetHmetArgMet----- 52
175 GTGAGAACCTTCGCTCAACAGAGCCCTGACCCCACTGCTCCGTATCGCTGTCCTC 234
53 ---SerTyrGlySerIlelysglyTyrgluProGlnAspGlyAlaTrrpYrsAntyrHis 71
235 GACTCCTCGAGCAGCAGCAGCTAGCGAGAGCCTTCTCC-----TATGGCTATGGA 285
72 ThrThrglyLysGlyValleuGluysGlnAspProlyssSerAspGluPheAlaValGln 91
286 GGCACCTGGAG-----AACGCCACCAAT----- 309
92 GlnAsnIleLeuAspLeuPheArgrThrylsAsnTyrgly---ArgTyrAlaGluAsnGly 110
310 -----AGCCGTTTGAACCTACGGAGACAGATTTCACAG----- 345
111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrgly----- 124
346 -----AACGATGTCATTCGCTGCTTCGCGATTTCGAA 378

QY 125 ---GlyAsnSerGlySerProValPheAspIysAsnGlyArgLeuIleGlyLeuAlaPhe 143
|||||
Db 379 TGTGAAATGACGTGACACTGCTTTTACCAAGATGAAATGATGGCATTCCTTC 438
QY 144 AspGlyAsnTPGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 439 CGAATCCAGAAAGAGCCCTGTATCTCATCTCATGTCTGCTGTAAGAAAT 498
QY 164 IleSerValAspIleArgTyr 170
Db 499 TCCGACGTGCACTTCACACTTC 519
RESULT 8
BQ215538 852 bp mRNA linear EST 02-MAY-2002
LOCUS AGENCOURT.7566326 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6043333
DEFINITION 5', mRNA sequence.
ACCESSION BQ215538
VERSION BQ215538.1 GI:20396938
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 852)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LLM13284 row: c column: 14
High quality sequence stop: 673.
Location/Qualifiers
1..852
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6043333"
/clone_lib="NIH_MGC_92"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: testis; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.5 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 198 a 228 c 239 g 186 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 0.372 Length: 852
Score: 89.50 Matches: 42
Percent Similarity: 38.32% Conservative: 22
Best Local Similarity: 25.15% Mismatches: 56
Query Match: 8.92% Indels: 47
DB: 14 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x BQ215538 (1-852)
QY 19 AlATrYrAlaIleGluLySGlyLysArgLeuPheAlaGlyLeuArgGluMetYrPro 38
Db 103 AGCTATGGGCTGAGAGGCCGCTGTATGCTCGAGATGAAATCAATGAGGAATCTCC 162
QY 39 GYAAGAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52
Db 163 GTGAAGACACCTTCCTCTACAGAGCCTGACCCCAAGCTGCTGATGCGCTGTCCTCG 222

QY 53 ---SerTyrGlySerIleLeuGlyTyrGluProGlnAspGlyAlaTrpTyrAsnTyrHis 71
|||||
Db 223 GACTCTGTGAGCACCAGCAGTACGACAGACCTTCTCCTC-----TATGCTATGGA 273
QY 72 ThrTrpGlyLySGlyValLeuGluIuLyGlnAspProLySerAspGluPheAlaValGln 91
Db 274 GGCACCTGGGAG-----AAGTCCACCAAT----- 297
QY 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGly---ArgTyrAlaGluAsnGly 110
Db 298 -----AGCCGCTTTGAAACTACGAGACAACTTTGCAAG----- 333
QY 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrGly----- 124
Db 334 -----AACGATGTGATTTGCTGCTTTCGCGATTTTGA 366
QY 125 ---GlyAsnSerGlySerProValPheAspIysAsnGlyArgLeuIleGlyLeuAlaPhe 143
Db 367 TGTGAAATGACGTGACACTGCTTTTACCAAGATGAAATGATGGCATTCCTTC 426
QY 144 AspGlyAsnTPGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 427 CGAATCCAGAAAGAGCCCTGTATCTCATCTCATGTCTGTAAGAAAT 486
QY 164 IleSerValAspIleArgTyr 170
Db 487 TCCGACGTGCACTTCACACTTC 507
RESULT 9
BE799506 853 bp mRNA linear EST 20-SEP-2000
LOCUS 601589315P1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3943490 5',
DEFINITION mRNA sequence.
ACCESSION BE799506
VERSION BE799506.1 GI:10220704
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 853)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DRP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at: image.llnl.gov
Plate: LUCM798 row: n column: 03
High quality sequence start: 24
High quality sequence stop: 790.
Location/Qualifiers
1..853
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3943490"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 202 a 203 c 255 g 193 t

ORIGIN

Alignment Scores:			
Pred. No.:	0.373	Length:	853
Score:	89.50	Matches:	42
Percent Similarity:	38.32%	Conservative:	22
Best Local Similarity:	25.15%	Mismatches:	56
Query Match:	8.92%	Indels:	47
DB:	12	Gaps:	8

US-10-008-355-2_COPY_522_712 (1-191) x BEF99506 (1-853)	
OY 19 AAlATyRAlAlleGluLySGlyLysArgLeuPhePheAlaGlyLeuAryGlumEtyrPro 38	
Db 288 AGCTATGGGCGTCAGAAAGGGCCGTATGCCTTCAGATGAATCAAGAAGMAATCTCC 347	
OY 39 GlytHAgAlaLeuProSerAspAla-----AsnPhetHmEtaArgMet----- 52	
Db 348 GTGAAGACACTTCCGCTCTACAGAGCCTACCCCACGCTGCTCATATGCGTGCTCCTG 407	
OY 53 ---SertyrGlySerIleLeuGlyTyrgLuProGlnAspGlyAlaTrpyrAsnTyrlHis 71	
Db 408 GACTCTCTCACACCACCACTAGCGAAGAGCCTTTCTCC-----TAGGCTATGGA 458	
OY 72 ThrtHrGlyLysGlyValLeuCluLysLnsPrPolySerAspGluPheAlaValGln 91	
Db 459 GGCACTGGGAAG-----AGTCACCAAT----- 482	
OY 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrgLy---ArgTYrAlaGluAsnGly 110	
Db 483 -----AGCCGGTTTGAANAACAGAGAACAGATTGCCAGAG----- 518	
OY 111 GluLeuHisIleAlaPheLeuSerAsnAsnAspIleThrgly----- 124	
Db 519 -----AACGATCGATTGGCTGCTGTTGCCGATTTTGAA 551	
OY 125 ---GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe 143	
Db 552 TGTGGAATGACGTGGGAAGCTGCTTTTACCAAGATTGGAAGTGATGGCATTCCTTTC 611	
OY 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163	
Db 612 CGAATCCAGAGGAAGACTTGCGGGGCTCAGGCCCTCTATCTCATGCTCGTGTGAAGAT 671	
OY 164 IleSerValAspIleArgTyrr 170	
Db 672 TGCCGAGTGGAGTTCACACTTC 692	

RESULT 10	881 bp	mRNA	linear	EST 05-FEB-2002
BM456284				
LOCUS				
DEFINITION	AGENCOURT_6409930 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:549635			
ACCESSION	BM456284			
VERSION	BM456284.1 GI:18505324			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (bases 1 to 881)			
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.			
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)			
	Unpublished (1999)			
	Contact: Robert Strausberg, Ph.D.			
	Email: rgapbs@fmail.nih.gov			
	Tissue Procurement: Lou Staudt			
	cDNA Library Preparation: Life Technologies, Inc.			
	CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)			
	DNA sequencing by: Agencourt Bioscience Corporation			
	Clone distribution: MCC clone distribution information can be			
	found through the I.M.A.G.E. Consortium/LLNL at:			
	http://image.llnl.gov			

Plate: LLAM12125 row: h column: 12
High quality sequence stop: 664.
Location/Qualifiers

FEATURES	Location/Qualifiers
source	1. .881

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5496635"
/clone_id="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph, Vector: PCMV-SPORT6; Site: 1: NC1;
Site: 2: Salt; Cloned unidirectionally; oligo- or plimded.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies
Note: this is a NIH_MGC library."

```

BASE COUNT	207 a	224 c	256 g	192 t	2 others
ORIGIN					

Alignment Scores:

Pred. No.:	0.393	Length:	881
Score:	89.59	Matches:	42
Percent Similarity:	38.32%	Conservative:	22
Best Local Similarity:	25.15%	Mismatches:	56
Query Match:	8.92%	Indels:	47
DB:	13	Gaps:	8

US-10-008-355-2_COPY_522_712 (1-191) X BM456284 (1-881)

OY		19	AlatyrAlaIleLeuGlyLysArgLeuPheAlaGlyLeuArgGluMetThrPro	38
Db		273	AGCTATGGGTGCACAGAAGGCCGCTGTATGCTTCGAGATGAAGATCAATCAGAAATCTCC	332
OY		39	GlyArGaLaLeuProSerAspAla-----AsnPherHrMetArgMet-----	52
Db		333	GTGAAGCACCTTCCGTCTACAGAGCCGTCGACCCTCCACTCGTGTCCTGATGGCTGATCCCTG	392
OY		53	---SerTrgLySerIleLeuSGLYTrgLuProGlnAspGlyAlaTrpTyraSnrHis	71
Db		393	GACTCTCCACAGACCACCTAGCTAAGCGAAGAGCCTTTCTCC-----TAGGCTATGCA	443
OY		72	ThrTrgGLyLySGlyValLeuGluLysGlnAspProLysSerAspGluPheAlaValGln	91
Db		444	GGCACTGGGAG-----AACTCCACCATT-----	467
OY		92	GluAsnIleLeuAspLeuPheArgTrhLysAsnTrgLy---ArgTYrAlaGluAsnGly	110
Db		468	-----AGCCGGTTTGAATAACAACGAGACAACTTGTCAGAG-----	503
OY		111	GlnLeuHisIleAlaPheLeuSerAsnaSnAspIleThrGly-----	124
Db		504	-----AACATGTGATTTGGCTGCTTGGCGATTTTCAA	536
OY		125	--GlyAsnSerGlySerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPhe	143
Db		537	TGTGGAAATGACGCGGAAGCTGCTCTTTACCAAGAATGGAAGTGATGGCATTCCTTTC	596
OY		144	AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr	163
Db		597	CGAATCCAGAAAGAAAGCTTGTGGGGGTGAGGCCCTCTATCCTCATGTCTGGTGAAGAT	656
OY		164	IleSerValaSpIleArgTyr	170
Db		657	TGCGGAGTGGAGTCAACTTC	677
RESULT	11			
LOCUS	BM449938			
DEFINITION	BM449938	899 bp	mRNA	linear EST 05-FEB-2002
ACCESSION	AGNCOCUFT_6392962	NIH_MGC_72	Homo sapiens	cDNA clone IMAGE:5528678
VERSION	BM449938			
KEYWORDS	EST.	BM449938.1	GI:18498978	
SOURCE	human.			

[illegible]

Db	426	CGAATTCAGAGAGAGACCTTGGGGGGTACAGGCCCTCTATCTCTCATGTCGTCTGTAAGAAAT	485
Qy	164	IIeSerValAspIleArgTyr	170
		::: :::	
Db	486	TGCCGAGTGGAGTTCACTTC	506
RESULT 12			
LOCUS	B0645307	904 bp	mRNA
DEFINITION	AGENCOURT_8302363 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:6271209		
ACCESSION	B0645307		
VERSION	B0645307.1	GI:21769479	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 904)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/ .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: CGAP (Stanford)		
	cDNA Library Preparation: Rubin Laboratory		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Agencourt Bioscience Corporation		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov		
	Plate: L10CM2447	row: j	column: 10
	High quality sequence stop: 741.		
FEATURES	Location/Qualifiers		
Source	1..904		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:6271209"		
	/clone_1id="NIH_MGC_100"		
	/tissue_type="hepatocellular carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: liver; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGCAGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH-MGC library."		
BASE COUNT	217 a	227 c	259 g
ORIGIN			201 t
Alignment Scores:			
Pred. NO.:	0.411	Length:	904
Score:	89.50	Matches:	42
Best Similarity:	38.32%	Conservative:	22
Percent Similarity:	25.15%	Mismatches:	56
Query Match:	8.92%	Indels:	47
DB:	14	Gaps:	8
US-10-008-355-2_COPY_522_712 (1-191) x B0645307 (1-904)			
Qy	19	AlaTyrAlaIleGluLyGlyLysArgGluPhePheAlaGlyLeuArgGluMetTyrPro	38
	::: :::		
Db	239	AGCTATGGGGTCAGAGGGCGCGTATGCTTCAGATGAAGATCAATGAGGAATCTCC	298
	::: :::		
Qy	39	GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet-----	52
	:::		
Db	299	GTCAGACACCTTCCTCTACAGAGCCTGACCCCAAGTGGTCCGTATCGCTGCTCCTG	358
	:::		
Qy	53	---SerTyrGlySerIleLeuGlyTyrGluProGlnAspGlyAlaTyrTyrAsnTyrHis	71
	::: :::		

Db 359 GACTCCTGACGACCACCGACGAGAGCCTTCTCC-----TATGGCTATGGA 409
Qy 72 ThrThrglyLysglyValleuGluysGlnAspProlySerAspLupheAlaValGln 91
Db 410 GGCACTGGGAG-----AAGTCACCAAT----- 433
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrgly---ArgTyraLagluAsngly 110
Db 434 -----AGCCGGTTGAAACTACGAGACAGTTCACAG----- 469
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrgly----- 124
Db 470 -----AACGATGTGATTGCTGCTTCGCGATTTCGA 502
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnglyArgLeuIleGlyLeuAlaPhe 143
Db 503 TGTGGAATGACGTGACACTGCTTTACCAAGAAATGAAAGATGGGCAATTCCTTTC 562
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 563 CGAATCCAGAGAGAACCTTGGGGGCTCAGGCCCTCTATCCTCATCTCCTGCGTGAAGAT 622
Qy 164 IleSerValAspIleArgTyr 170
Db 623 TGGCAGTGGAGTTCAACTTC 643

RESULT 13
LOCUS B0652808 942 bp mRNA linear EST 15-JUL-2002
DEFINITION AGENCOURT_8490114 NIH_MGC_100 Homo sapiens CDNA clone IMAGE:6298491
ACCESSION B0652808
VERSION B0652808
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS NIH-MGC
TITLE NIH-MGC http://mnc.nci.nih.gov/ (bases 1 to 942)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2509 row: k column: 04
High quality sequence stop: 621.

FEATURES
source
Location/Qualifiers
1. 942

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6298491"
/clone_lib="NIH_MGC_100"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pOTB7; Site: 1; XhoI; Site: 2;
ecori; CDNA made by oligo-dt priming. Directionally cloned
into EcorI/XhoI sites using the following 5' adaptor:
GGCAGCAGC(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 222 a 240 c 275 g 205 t
ORIGIN

Alignment Scores:

Pred. No.:	0.44	Length:	942
Score:	89.50	Matches:	42
Percent Similarity:	38.328	Conservative:	22
Best Local Similarity:	25.15%	Mismatches:	56
Query Match:	8.92%	Indels:	47
DB:	14	Gaps:	8

US-10-008-355-2_COPY_522_712 (1-191) x B0652808 (1-942)

Qy 19 AlaTyraIleGluLysGlyArgLeuPheAlaGlyLeuArgLumetyrPro 38
Db 239 AGCTATGGGGTGAGAGGGGCCCTGATGCTTCGACATGAAATCAATGGAATCTCC 298
Qy 39 GlyArgAlaLeuProSerAspAla-----AsnPheThrMetArgMet----- 52
Db 299 GTGAAACACCTTCCTCCGCTACAGACCTGACCCCAAGTGGTCCGTATCGGTCCTCG 358
Qy 53 ---SerTyrglySerIleLysGlyTyrgluProGlnAspGlyAlaTrpTyraAsnTyrlHis 71
Db 359 GACTCCTGACGACCCAGCTAGCGAGAGACCTTCTCC-----TATGGCTATGGA 409
Qy 72 ThrThrglyLysglyValleuGluysGlnAspProlySerAspLupheAlaValGln 91
Db 410 GGCACTGGGAG-----AAGTCACCAAT----- 433
Qy 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrgly---ArgTyraLagluAsngly 110
Db 434 -----AGCCGGTTGAAACTACGAGACAGTTCACAG----- 469
Qy 111 GlnLeuHisIleAlaPheLeuSerAsnAspIleThrgly----- 124
Db 470 -----AACGATGTGATTGCTGCTTCGCGATTTCGA 502
Qy 125 ---GlyAsnSerGlySerProValPheAspLysAsnglyArgLeuIleGlyLeuAlaPhe 143
Db 503 TGTGGAATGACGTGACACTGCTTTACCAAGAAATGAAAGATGGGCAATTCCTTTC 562
Qy 144 AspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
Db 563 CGAATCCAGAGAGAACCTTGGGGGCTCAGGCCCTCTATCCTCATCTCCTGCGTGAAGAT 622
Qy 164 IleSerValAspIleArgTyr 170
Db 623 TGGCAGTGGAGTTCAACTTC 643

RESULT 14
LOCUS BG174574 943 bp mRNA linear EST 06-FEB-2001
DEFINITION 602334417F1 NCI_CGAP_Mam1 Mus musculus CDNA clone IMAGE:4457381 5',
ACCESSION BG174574
VERSION BG174574.1 GI:12681277
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS NIH-MGC
TITLE NIH-MGC http://mnc.nci.nih.gov/ (bases 1 to 943)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM10254 row: b column: 06
High quality sequence stop: 656.

Db 503 TGTGAATGACGTGGAACGTCTTTTACCAAGATGGAAGTGGATGGCATTCCTTTC 562
 QY 144 AspGlyAsnTPGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThr 163
 Db 563 CGAATCCAGAAAGAGCCTTGGGGGGTCAGGCCCTCTATCCTCATGTCTCGTGGAAGAT 622
 QY 164 IleSerValAspIleArgTyr 170
 Db 623 TGGCGAGTGGAGTTCACTTC 643

Search completed: December 21, 2002, 11:19:49
 Job time : 2176 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: December 21, 2002, 01:47:47 : Search time 59 seconds

(without alignments)
1284.742 Million cell updates/sec

Title: US-10-008-355-2_COPY-522_712

Sequence: 1 SKSVIAARAIQADAMANY.....LFMIDKMGCCPRLEIKI 191

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 356696 seqs, 198428768 residues

Total number of hits satisfying chosen parameters: 713392

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame.p2n.model -DEV=x1h
-Q/cg2_1/USPTO.spool/US10008355/runat_11722002_112705_17000/app.query.fasta_1.327
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100
-THR.MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pro -NORM=ext -HEARST=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10008355 -CGCN_1_1_21_@runat_11722002_112705_17000
-NCPU=6 -ICPU=3 -NO.XLPHY -NO.MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA.*

1: /cg2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cg2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq.*
3: /cg2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cg2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cg2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cg2_6/ptodata/2/pubpna/PTCTUS_PUBCOMB.seq.*
7: /cg2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cg2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cg2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
10: /cg2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cg2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
12: /cg2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
13: /cg2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
14: /cg2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1003	100.0	2139	9	US-10-008-355-1	Sequence 1, Appl1
2	85.5	8.5	651	10	US-09-770-149-525	Sequence 525, App
3	81.5	8.1	720	10	US-09-815-242-4296	Sequence 4296, Ap
4	81.5	8.1	765	10	US-09-815-242-8179	Sequence 8179, Ap

5	81.5	8.1	1812	9	US-09-938-842A-2020	Sequence 2020, Ap
6	81	8.1	849	12	US-10-062-624-3	Sequence 3, Appl1
7	79.5	7.9	861	10	US-09-974-300-2760	Sequence 2760, Ap
8	77.5	7.7	841	9	US-09-996-634-39	Sequence 39, Appl
9	76.5	7.6	1026	10	US-09-815-242-7701	Sequence 7701, Ap
10	76	7.6	888	10	US-09-974-300-1729	Sequence 1729, Ap
11	75	7.5	1986	10	US-09-974-300-1841	Sequence 1841, Ap
12	74.5	7.4	1026	10	US-09-991-258-9	Sequence 9, Appl1
13	74.5	7.4	5076	10	US-09-952-360-8	Sequence 8, Appl1
14	74.5	7.4	14759	10	US-09-952-360-1	Sequence 1, Appl1
15	73.5	7.3	1472	10	US-09-974-300-1065	Sequence 1065, Ap
16	73	7.3	1282	9	US-09-987-107-43	Sequence 43, Appl
17	73	7.3	1311	10	US-09-815-242-7057	Sequence 7057, Ap
18	73	7.3	3170	9	US-10-135-322-4	Sequence 4, Appl1
19	73	7.3	3170	9	US-10-135-322-23	Sequence 23, Appl
20	73	7.3	3170	9	US-09-918-508-5	Sequence 5, Appl1
21	73	7.3	3503	9	US-10-135-322-2	Sequence 2, Appl1
22	73	7.3	3612	9	US-10-135-322-3	Sequence 3, Appl1
23	73	7.3	3620	9	US-10-135-322-1	Sequence 1, Appl1
24	73	7.3	11517	10	US-09-901-106-1	Sequence 1, Appl1
25	73	7.3	70768	9	US-10-135-322-13	Sequence 13, Appl
26	72.5	7.2	842	12	US-10-062-994-3	Sequence 3, Appl1
27	72.5	7.2	1188	9	US-09-938-842A-615	Sequence 615, App
28	72.5	7.2	1285	9	US-09-987-107-45	Sequence 45, Appl
29	72.5	7.2	2895	10	US-09-998-598-369	Sequence 369, App
30	72.5	7.2	352	10	US-09-867-701-4791	Sequence 4791, Ap
31	72	7.2	852	12	US-10-059-964-39	Sequence 39, Appl
32	72	7.2	1305	9	US-09-938-842A-2603	Sequence 2603, Ap
33	72	7.2	2188	10	US-09-925-300-598	Sequence 598, App
34	72	7.2	1996	10	US-09-974-300-6219	Sequence 6219, Ap
35	71.5	7.1	996	10	US-09-815-242-9413	Sequence 9413, Ap
36	70	7.0	1704	10	US-09-757-781-20	Sequence 20, Appl
37	69.5	6.9	824	10	US-09-956-004-76	Sequence 76, Appl
38	69.5	6.9	2847	9	US-09-938-842A-632	Sequence 632, App
39	69.5	6.9	9775	10	US-09-751-962-1	Sequence 1, Appl1
40	69.5	6.9	1137	9	US-09-938-842A-1746	Sequence 1746, Ap
41	69	6.9	1299	10	US-09-815-242-6177	Sequence 6177, Ap
42	69	6.9	22960	10	US-09-070-927A-345	Sequence 345, App
43	68.5	6.8	2900	12	US-10-044-090-270	Sequence 270, App
44	68.5	6.8	2921	10	US-09-880-107-3319	Sequence 3319, Ap
45	68.5	6.8				

ALIGNMENTS

RESULT 1
US-10-008-355-1
; Sequence 1, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Novartis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235 00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-1

Alignment Scores:
Pred. No.: 1.03e-123
Score: 1003.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
Length: 2139
Matches: 191
Conservative: 0
Mismatch: 0
Indels: 0

DB:	9	Gaps:	0
US-10-008-355-2_COPY_522_712 (1-191) x US-10-008-355-1 (1-2139)			
Qy 1 SerlyssSerValIlealaalaAaArgAlaIleGlnAlaaspAlaamelaAaanaIaTyr 20			
Db 1564 TCCAAGAGCGTAATAGCTGCTGCGCGGCGATTCACAGCGCGATGCGATGGCCAAATCCCTAT 16233			
Qy 21 AlaIleGluYsgIyLysAqGleuPhePheAlaGlyLeuArgGluMetYrProGlyArg 40			
Db 1624 GCCATTGGAAGGGCGCAAGCCCTCTTTCTTTGCGGTTCCGTGAGATGTACCCCGGAGCT 16833			
Qy 41 AlaIeuProSerAspAlaAaenPhePheMetArgMetSerTyrGlySerIleYsgIyTyr 60			
Db 1664 GCTGTCGCCAGCGATGGCCAACTTCACCATCGATAGCTAGCGCTCCATCAAGGGATAT 17433			
Qy 61 GluProGlnAspGlyAlaTyrPtyAsnTyrHisThrThrGlyLysGlyValLeuGluLys 80			
Db 1744 GAACCGCAGAGCGGTGGCTGGTGCATCATACACACAGCGCATATTTGGAGAG 18033			
Qy 81 GlnAspProLysSerAspGluPheAlaValGlnGluAsnIleLeuAspLeuPheArgThr 100			
Db 1804 CAGATCCTAAGACGCATGATGATTGGCCGTCACAGAGAAATATCCTGCACCTTCCGCAC 18633			
Qy 101 LysAsnTyrGlyArgTyrAlaGluAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 120			
Db 1864 AAAAACAATAGTCCCTVTTGGCCGGAACGGTCACCTCATATGCTTCTCATCCGAACAC 19233			
Qy 121 AspIleThrGlyGlyAsnSerGlySerProValPheAspLysAsnGlyValArgLeuIleGly 140			
Db 1924 GACATCACGGCGGTAACTCCGTAACCCCGTATTGATATAACAGCGCCGTTGATCGGT 19633			
Qy 141 LeuAlaPheAspGlyAsnTyrPgluAlaMetSerGlyAspIleGluPheGluProAspLeu 160			
Db 1984 CTTCCTTCATGATGCACACTGGGAAGCATATGAGTGGTGCATCATGAGTGAACCCGATCTG 20433			
Qy 161 GlnArgThrIleSerValAspIleArgTyrValLeuPheMetIleAspLysTyrPglu 180			
Db 2044 CAGGCGCAATACACCGTGGACATCCGCTACGTCTCTCATGATATGACAAATGGGGTACG 21033			
Qy 181 CysProArgLeuIleGlnGluLeuLysLeuIle 191			
Db 2104 TGCCCCCGCTCTCATCCAAAGCTGAAAGTTGATC 2136			
RESULT 2			
US-09-770-149-525			
Sequence 525, Application US/09770149			
Patent No. US20020059663A1			
GENERAL INFORMATION:			
APPLICANT: Gorlach, Jorn			
APPLICANT: An, Yong-Qiang			
APPLICANT: Hamilton, Carol M.			
APPLICANT: Price, Jennifer L.			
APPLICANT: Raines, Tracy M.			
APPLICANT: Yu, Yang			
APPLICANT: Rameeka, Joshua G.			
APPLICANT: Page, Amy			
APPLICANT: Matthew, Abraham V.			
APPLICANT: Ledford, Brooke L.			
APPLICANT: Moessner, Jeffrey P.			
APPLICANT: Haas, William David			
APPLICANT: Garcia, Carlos A.			
APPLICANT: Krickler, Maja			
APPLICANT: Slader, Ted			
APPLICANT: Davis, Keith R.			
APPLICANT: Allen, Keith			
APPLICANT: Hoffman, Neil			
APPLICANT: Hurban, Patrick			
TITLE OF INVENTION: Expressed Sequences of Arabidopsis			
FILE REFERENCE: 2024 (PARR-013PRV)			
CURRENT APPLICATION NUMBER: US/09770,149			
CURRENT FILING DATE: 2001-01-26			

```

PRIOR APPLICATION NUMBER: 60/178,506
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 525
LENGTH: 651
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(651)
OTHER INFORMATION: n = A,T,C or G
US-09-770-149-525

Alignment Scores:
Pred. NO.: 0.0206 Length: 651
Score: 85.50 Matches: 42
Percent Similarity: 38.16% Conservative: 16
Best Local Similarity: 27.63% Mismatches: 67
Query Match: 8.52% Indels: 27
DB: Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x US-09-770-149-525 (1-651)
OY 45 AspAlaAsnPhetThrMetAlaGMetSerTyrGlySerIleGlyGlyPro--Gln 63
Db 72 GATGCTTAAGGGAGACAGATTATTCAGCAAGGAAATGTGGCTTGTATCCAGATAT 1311
OY 64 AspGlyAlaIlePrrTyAsnTyrHisThtThGlyLys-----Gly 76
Db 132 GATCTAGCTGTGTTGTAAGATTTGAACCTGAGGAGGAGGCTGACTAAATCCTGTTCTCGGT 1911
OY 77 ValLeuGluLysGlnAspProLysSerAspLeuPheAlaVal-----Gln 91
Db 192 ACCTCCATGATGATACCGGGTAGTCGCAAAAGTTCGATTTGGATTCATATGATAT 2511
OY 92 GluAsnIleLeuAspLeuPheArgThrLysAsnTyrGlyArgTyr-----AlaGluAsn 1099
Db 252 GAAAAACACCTTGTACAAATAGGGGTAGTAGGGCTTGGGAGAGAGATACCTTCACCTAAT 3111
OY 110 GlyGln---LeuHisIleAlaPheLeuSerAsnAspIleThrGlyGlyAsnSerGly 1288
Db 312 GGGAGAGACTATTTCAGAGACTATATCAAAACCGATCTGTATTTAACTCAGGCAATTTGGA 3711
OY 129 SerProValPheAspLysAsnGlyArgLeuIleGlyLeu-----AlaPheAspGly 1455
Db 372 GGGGCATTGCTGTGATTCTTATAGGCCATACCATAGTGTGAACAACGTGCACATTCACCGA 4311
OY 146 AsnTrpGluAlaMetSerGlyAspIleGluPheGluProAspLeuGlnArgThrIleSer 1655
Db 432 AAAGGAGTGTGATGTCTTCTTGGTGTAAACTTT-----GCCATTCCC 4731
OY 166 ValAsp-----IleArgTyrValLeuPheMetIle 175
Db 474 ATTGACACAGTTGTCCGAACAGTTCGCCATCCTCAATT 509

RESULT 3
US-09-815-242-4296
Sequence 4296, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlssen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242

```

```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4296
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4296

Alignment Scores:
Pred. No.: 0.031 Length: 720
Score: 81.50 Matches: 34
Percent Similarity: 37.34% Conservative: 22
Best Local Similarity: 22.97% Mismatches: 59
Query Match: 8.13% Indels: 33
DB: 10 Gaps: 6

US-10-008-355-2_COPY_522_712 (1-191) x US-09-815-242-4296 (1-720)
QY 37 TYRPROGLYARGALALEUPROSERASPALAASNPHERHMETARGMETSERTYRGLYSER 56
DB 322 TATCCAGGTAAGAAGATGATCATTCATCAAGTTGAAGCGCAGCAATAGACGTGCT 381
QY 57 ILEYSGLYTYGULUPROGLINASPGLYALATRP--TYRASNTRYHISTHRTHRGLYLS 75
DB 382 CCAAAAGCATTTAATTATATGATATGTAACGCCGTTAAATATGACACAGGGCTTAA 441
QY 76 GLY-----ValleugluLysGlnAspProLysSerAspLuphealVal 90
DB 442 GCTGTCACCGAATTAAGTATCGTGTATCCACACCCATACAAAATAATATGTTTA 501
QY 91 GINGLUASNILLEUASPLEUPHEARGTHRLYSASN-----TYRGLYARGTYR 106
DB 502 CATGAGTCAACTGCTCGTGTATGTCAGTAGAAGTAGCAGCATTTGTATATTCAGCGCAT 561
QY 107 ALAGUASNGLYGLINLEUHNISILEALAPHELEUSERASNASNAPLIERHRTGLYLYASN 126
DB 562 ACTGAAGT-----GGAAGC 576
QY 127 SERGLYSERPROVALPHEASPLYSASNGLYARGLEUILEGYLEUALAPHE----- 143
DB 577 TCTGATACCTCGTATTAACAGCAACGAAATTAATAGTATTCTTTGCTTCTGAT 636
QY 144 -----AspGLYASNTRYPGIUALAMETSERGLYASPILEGLUPHEGLUPROASPLEU 160
DB 637 GTAAAAAATGATGACACAGAAATGCATATGCG-----GCTACTTTACACGAGAAAT 690
QY 161 GINARGTHRILLESERVALASPILE 168
DB 691 AAAAAGTTCATTGCAAGAAACATA 714

RESULT 4
US-09-815-242-8179
; Sequence 8179, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
```

```

; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8179
; LENGTH: 765
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(765)
US-09-815-242-8179

Alignment Scores:
Pred. No.: 0.0883 Length: 765
Score: 81.50 Matches: 34
Percent Similarity: 37.84% Conservative: 22
Best Local Similarity: 22.97% Mismatches: 59
Query Match: 8.13% Indels: 33
DB: 10 Gaps: 6

US-10-008-355-2_COPY_522_712 (1-191) x US-09-815-242-8179 (1-765)
QY 37 TYRPROGLYARGALALEUPROSERASPALAASNPHERHMETARGMETSERTYRGLYSER 56
DB 364 TATCCAGGTAAGAAGATGATCATTCATCAAGTTGAAGCGCAGCAATAGACGTGCT 423
QY 57 ILEYSGLYTYGULUPROGLINASPGLYALATRP--TYRASNTRYHISTHRTHRGLYLS 75
DB 424 CCAAAAGCATTTAATTATATGATATGTAACGCCGTTAAATATGACACAGGGCTTAA 483
QY 76 GLY-----ValleugluLysGlnAspProLysSerAspLuphealVal 90
DB 484 GCTGTCACCGAATTAAGTATCGTGTATCCACACCCATACAAAATAATATGTTTA 543
QY 91 GINGLUASNILLEUASPLEUPHEARGTHRLYSASN-----TYRGLYARGTYR 106
DB 544 CATGAGTCAACTGCTCGTGTATGTCAGTAGAAGTAGCAGCATTTGTATATTCAGCGCAT 603
QY 107 ALAGUASNGLYGLINLEUHNISILEALAPHELEUSERASNASNAPLIERHRTGLYLYASN 126
DB 604 ACTGAAGT-----GGAAGC 618
QY 127 SERGLYSERPROVALPHEASPLYSASNGLYARGLEUILEGYLEUALAPHE----- 143
DB 619 TCTGATACCTCGTATTAACAGCAACGAAATTAATAGTATTCTTTGCTTCTGAT 678
QY 144 -----AspGLYASNTRYPGIUALAMETSERGLYASPILEGLUPHEGLUPROASPLEU 160
DB 679 GTAAAAAATGATGACACAGAAATGCATATGCG-----GCTACTTTACACGAGAAAT 732
```

```

Qy 161 GluArgThrIleSerValaspIle 168
Db 733 AAAAGTTCATTGCAGAAAACATA 756

RESULT 5
US-09-938-842A-2020
; Sequence 2020, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; FILE REFERENCE: S00-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2020
; LENGTH: 1812
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2020

Alignment Scores:
Pred. No.: 0.299 Length: 1812
Score: 81.50 Matches: 53
Percent Similarity: 38.05% Conservative: 33
Best Local Similarity: 23.45% Mismatches: 67
Query Match: 8.13% Indels: 74
DB: Gaps: 13

US-10-008-355-2_COPY_522_712 (1-191) x US-09-938-842A-2020 (1-1812)
Qy 18 AsnAlaItyr-----AlaIleGluysGlyArgLeuPheAlaIleuArg--- 34
Db 421 AATCTTACGTCGCGGATAGACAAAGGATTACTCTTATTCGTGAAGTCG 480
Qy 35 GluMetTyProGlyArgAlaLeuProSerAspAlaAsnPhetHr-----MetArgMet 52
Db 481 GAGTCAAGACGCTCAAGTGGCTTACCGGCTCGCTCACCAGATTATACAAAGC 540
Qy 53 SerTyrgIySerIleIysGlyTyrgIuProGlnAspGlyAlaItrTyAsnTyHstHr 72
Db 541 GACCACTTCAGAGACAGACCGTACGATCG-----TACAACTTCACACT 585
Qy 73 ThrGlyIysGlyValLeuGluIysGlnAspProIysSerAspGluPheIleA----- 89
Db 586 ACTCCTTAAGCAAGCACTCTGCTCGCTCCGCCAAGCATGATGCTCAAAATGCTA 645
Qy 90 -----ValGlnGluAsnIleLeuAspLeuPheArgThrIlysAsnTyrgIyArg 105
Db 646 TGTGCTCTCTTAATGCCCATGAAGTTCTTCGCATC-----GGCGCA 687
Qy 106 TyraIaGluAsnGlyGlnLeuHis---IleAlaPheLeuSerAsnAsn----- 120
Db 688 GTGTTTGCTTCGCGTCTCTCGCTCGCATTAAGCTTCCTCCAGAACAAATGGAAGCACT 747
Qy 121 -----AspIleThrGlyIysAsnSerGlySerProValPheAsp-----LysAsn 135
Db 748 GCTCGATATCTCAACCGGAACCTTAAGTCTCGAATCTTGATCTGCACTTAAAC 807
Qy 136 GlyArg-----LeuIleGlyLeu 141
Db 808 CCGATGTCACAGATTTTGACCAAACTGATCAAGACGTGAGTTTGGTTGGGGTT 867

```

```

Qy 142 AlaPheAspGlyAsnTrpGluAlaMetSerGlyaspIle----- 154
Db 868 TGTTCACAGAGAAATTCGGAAGGATTAATCAACAAGATGGCTTAACAGAACTATC 927
Qy 155 -----GluPheGluProAspLeuGlnArgThrIleSer 165
Db 928 GACGTGATGTACTGAGCAATGCTCAGTATATCCGACCTTGAGT--ACTATAGCG 985
Qy 166 ValAspIleArgTyValLeuPheMetIleAspLysTrpGlyGlnCys-----Pro 182
Db 986 GTGATTTACCGATGGCTT-----GCACGATGTATGCTGCTCG 1024
Qy 183 ArgLeuIleGlnGluLeu 188
Db 1025 AAAGTATTTCGCGATT 1042

RESULT 6
US-10-062-624-3
; Sequence 3, Application US/10062624
; Patent No. US20020115840A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2/D1
; CURRENT APPLICATION NUMBER: US/10/062,624
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/660,587
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 3
; LENGTH: 849
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; NAME/KEY: mat_peptide
; OTHER INFORMATION: nucleic acid sequence of p28-5
US-10-062-624-3

Alignment Scores:
Pred. No.: 0.119 Length: 849
Score: 81.00 Matches: 43
Percent Similarity: 39.38% Conservative: 20
Best Local Similarity: 26.88% Mismatches: 67
Query Match: 8.08% Indels: 30
DB: Gaps: 7

US-10-008-355-2_COPY_522_712 (1-191) x US-10-062-624-3 (1-849)
Qy 23 GluIysGlyIysArgLeuPheAlaGlyLeuArgGluMetTyProGlyArgAlaLeu 42
Db 172 GAGAAAAAAGACACACTGATATATGCTTAAGAAAGAACTGGCGACGAGATGCAATA 231
Qy 43 ProSer-----AspAlaAsnPhetHrMetArg-----MetSerTyrgIySer 56
Db 232 TCTAGCAAAAGTCAGAGATATATTTTCCATTCGAATTAATCAATCAATGCAAGC 291
Qy 57 IleIys-----GlyTyrgIuProGlnAspGlyAlaItrTyAsnTyHstHrThrGly 74
Db 292 AACCAAGTTTATGGGTTTGCATGCTATTTGCT-----TACTCGATAGGC 336
Qy 75 LysGlyValLeuGluIysGlnAspProIysSerAspGluPheAlaValGlnGluAsnIle 94
Db 337 ACTCCAGCAATGAA-----GTTGAGATGCT 363
Qy 95 LeuAspLeuPheArgThrIlysAsnTyrgIyArgTyraIaGluAsnGlyGlnLeuHisIle 114
Db 364 TATGAAGCATTTGATGTGAATAATCCAGGTGATTAATTAACAAAGCGTGTACAGGTAT 423
Qy 115 AlaPheLeuSerAsnAsnIleThrGlyIysAsnSerGlySerPro----- 130

```

Db 424 TGTGCTTATCTCATCAATGATGCGGATGATGACATGACTAGTGCACATGCAAAATTT 483
Qy 131 VALPheasplYAsnGlyArgLeuileGlyLeuAlaPheaspGlyAsn-----TTPGlu 148
Db 484 GATATATTATTAATGAAGATTACTTAAACATATCATTTATGACAAAGATATGTTATGAA 543
Qy 149 AAmetSerGlyAspIleGluPheGluProaspLeuGlnArgThrIleSerValaSpile 168
Db 544 ACAGCAAGCAAAATATATCTCTCTCTCTTACATATGACAGGATTTGGTACTGATTTA 603
RESULT 7
US-09-974-300-2760
; Sequence 2760, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Id Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2760
; LENGTH: 861
; TYPE: DNA
; ORGANISM: *Bacillus licheniformis*
US-09-974-300-2760

Alignment Scores:
Pred. No.: 0.193 Length: 861
Score: 79.50 Matches: 37
Percent Similarity: 36.84% Conservative: 19
Best Local Similarity: 24.34% Mismatches: 39
Query Match: 7.93% Indels: 57
DB: 10 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x US-09-974-300-2760 (1-861)

Qy 43 PROSerAspAlaAsnPhenhrMetArgMetSerTyrGlySerIleGlySerGlyTyrGluPro 62
Db 455 CCGTGGATGCGCTTATCTATACGTTAGATTATGAGACA-----AAATCCGTT 505
Qy 63 GlnAspGlyAlaTyrPyrAsnTyrHisThrGlyLysGlyValLeuGluLysGlnAsp 82
Db 506 ACAAAACCGTGGATATGTTATGATATAAAAGGTTGAAGGCTATGTGAAGAAAGTCAGAC 565
Qy 83 Pro-----LysSerAspGluPheAlaValGlnLysGlnAsnIleLeu--- 95
Db 566 GCGGTTTGCACAAAGACCTTGAAACAGAAACCGGTTGTTCTTCAAAACATTCTAATA 625
Qy 96 -----AspLeuPheArg 99
Db 626 GCGGAAGCCGAACATCATGAAAGTCGTACGCGAGAGGAGGATATTACCTT---CGT 682
Qy 100 ThrIysAsnTyrGlyArgTyrAlaGlnAsnGlyGlnLeuHisIleAlaPheLeuSerAsn 119
Db 683 TCAGGGGGAAGAGACTCTCTGCAAAACGGGGAAGTCATCGGATCGAGCTGGAAGAAC 742
Qy 120 AsnAspIleThrGlyLysAsnSerGlySerProValPheAspLysAsnGlyArgLeuIle 139
Db 743 -----GAAACGCGCCGATCATATA 760
Qy 140 GlyLeuAlaPheaspGly-----AsnTrpGluAlaMet 150
Db 761 CCGAGCGCG-----GACGCGCAAAACCGTTCGCTTCGCGGGGAAAAACGTGG----- 808
Qy 151 SerGlyAspIleGluPheGluProaspLeuGlnArg 162

Db 809 -----ATCAATATCGTTCCCGATCTCAGCAAG 835
RESULT 8
US-09-996-634-39
; Sequence 39, Application US/09996634
; Patent No. US20020172684A1
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 61260
; CURRENT APPLICATION NUMBER: US/09/996,634
; CURRENT FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: 09/447,135
; PRIOR FILING DATE: 2000-01-03
; PRIOR APPLICATION NUMBER: 08/990,823
; PRIOR FILING DATE: 1997-12-15
; PRIOR APPLICATION NUMBER: US 96/10375
; PRIOR FILING DATE: 1996-06-14
; PRIOR APPLICATION NUMBER: 60/000,254
; PRIOR FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 169
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 39
; LENGTH: 841
; TYPE: DNA
; ORGANISM: *Mycobacterium tuberculosis*
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(841)
; OTHER INFORMATION: n is a, c, g, or t/u.
US-09-996-634-39

Alignment Scores:
Pred. No.: 0.344 Length: 841
Score: 77.50 Matches: 40
Percent Similarity: 32.91% Conservative: 12
Best Local Similarity: 25.32% Mismatches: 45
Query Match: 7.73% Indels: 61
DB: 9 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x US-09-996-634-39 (1-841)

Qy 36 MetTyrProGlyArgAlaLeuProSerAspAlaAsnPhenhrMetArg---MetSerTyr 54
Db 276 CTGTATCCG-----ACGGCAATGTGACCTACCGCGGTATCACCATT 317
Qy 55 GlySerIleLysGlyTyrGluProGlnAsp 64
Db 318 GGCAGAGTTACTCGCTGAGCCGACCCAGCCAGCGCAGAGTGCATGACATCGCC 377
Qy 65 -----GlyAlaTyrPyrAsnTyrHisThrGlyLysGlyVal 77
Db 378 AGCAACTACAAATATCCCGCGATGCGCTCGCGCAAGCTGATTCGGGTCA----- 428
Qy 78 LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnLysGlnAsnIleLeuAspLeu 97
Db 429 -----GCGGTGGCGAGCATCATCGACCTG 455
Qy 98 PheArgThrLysAsnTyrGlyArgTyrAlaGlnAsnGlyGlnLeuHisIleAlaPheLeu 117
Db 456 GTGTCCACCGGCTCTCCGGTAAATACTTCTCTCCGAGACG----- 497
Qy 118 SerAsnAsnAspIleThrGlyLysAsnSerGlySer-----ProValPheaspLys 134
Db 498 -----ACCATCACCAAGGCGACCGTTCCAGTGAAGATCGGGCGGCGGTGACAAAT 548
Qy 135 AsnGlyArg-----LeuIleGlyLeuAlaPheaspGlyAsn 146
Db 549 TCCCAATCGCGGGTGGCCCGCATTCGCCACGAGAAAGATGCGGCTTCGCTGCGTCGACGAGACC 608
Qy 147 TrpGluAlaMetSerGlyAspIleGluPheGluProaspLeuGlnArgThrIle 164


```
OY 75 sGlyVal-----LeuGluLysGlnAspProLysSerAspGluPheAlaValGlnG1 92
    |||||
Db 406 AGGTTCAAGACGCGCGGTGAGAGCATGATGCGGTAAACAGATGAA-----AACGG 456
OY 92 uAsnLleuAspLeuPheArgThrLysAsnTrpGlyArgTrpAlaGlnAsnGlyGlnLe 112
    |||||
Db 457 CAACTTGATCGGAAAAATCAAAAGACAAAGTGAACAATTTCACTTCCACCGCTCCAGCT 516
OY 112 uHisLleAlaPheLeuSerAsnAspLleThr-----GlyGlyAsnSerGlySer 129
    |||||
Db 517 TTTAAGCCGCAACAGACAGAAACACAGAAATCTCCACAGTATCGCTCGCATTAAGGCAA 576
OY 129 rProValPheAspLysAsnGlyArgLeuLleGlyLeuAlaPheAspGlyAsnTrpGluAl 149
    |||||
Db 577 GAAAAAGGCGT-----AACGGAATCATCAACGCG-----TACGACTCGATTAAGAAAGCG 624
OY 149 aMeterGlyAspLleGluPheGluProAspLeuGlnArgThrTle 164
    |||||
Db 625 GCTTTCATCGAATAATC--ATCGAGCCCTGATGAGACAGGAAGTG 667

RESULT 11
US-09-974-300-1841
; Sequence 1841, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085,500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1841
; LENGTH: 1986
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1841

Alignment Scores:
Pred. No.: 2,49 Length: 1986
Score: 75.00 Matches: 38
Percent Similarity: 35.47% Conservative: 23
Best Local Similarity: 22.09% Mismatches: 63
Query Match: 7.48% Indels: 48
DB: 10 Gaps: 8

US-10-008-355-2_COPY_522_712 (1-191) x US-09-974-300-1841 (1-1986)
OY 18 AsnAlaTrpAlaLleGluLysGlyArgLeuPheAlaGlyLeuArgGlu----- 35
    |||||
Db 271 AACGATCATCGTGAAGAAAGACTCGAGCG-----TACGCCAAGGTGAATCAGAGTGG 324
OY 36 -----MetTrpProGlyArgAlaLeuProSerAspAlaAsnPheThrMetArgMetSer 53
    |||||
Db 325 GAAGCGATCTACACAGGT-----TCAAAAAGCGCAAGTTCATCCGATATCCCGCC 375
OY 54 TrpGlySerLleLysGlyTrpGluProGlnAspGlyAlaTrpTrpAsnTrpThr 73
    |||||
Db 376 GAGAAATATGCGAAGAGGTATTAACCGGCGACAGGAGCTGTAATAGCAAGCGCTGAA 435
OY 74 GlyLysGly----- 76
    |||||
Db 436 AATTAAGAAAGGTGATCATCAAAACCTTATTAACAGCGTCAACCGAAGCATGCTC 495
OY 77 -----ValLeuGluLysGlnAspProLysSerAspGluPheAlaValGlnGlnLe 94
    |||||
Db 496 ATTACGATTGCGAAGAGACGAGGAGCGCTCAGAGATTGTGCGGTCAACATGAATAATC 555
```

```
OY 95 LeuAspLeuPheArgThrLysAsnTrpGlyArgTrpAlaGlnAsnGlyLnuLeuHisLle 114
    |||||
Db 556 GATGAGCTTATCAAGACGACAGAGCGCCGTCACATCTCGAAGAAAGCGCG-----TTT 606
OY 115 AlaPheLeuSerAsnAsp-----LleThrGlyGlyAsn 126
    |||||
Db 607 GCGTTTATTTCAGCAAGTACAAAAAATATGTGTGCCATCCGACATTCGAACCGGAAACC 666
OY 127 SerGlySerProValPheAspLysAsnGlyArgLeuLleGly----- 140
    |||||
Db 667 GAAGCGAAAGCG-----GACTGCGTTGGACAATGTACGAGTGAAGAAAGCGATTTT 720
OY 141 ---LeuAlaPheAspGlyAsnTrpGluAlaMetSer 151
    |||||
Db 721 CAATATACATTTCGAAGAAAGAAAGAAATGCGT 756

RESULT 12
US-09-991-258-9
; Sequence 9, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Robert
; APPLICANT: Keith, Paula
; APPLICANT: Dryga, Sergey
; APPLICANT: Caley, Ian
; APPLICANT: Maughan, Maureen
; APPLICANT: Johnston, Robert
; APPLICANT: Davis, Nancy
; APPLICANT: Swanson, Ronald
; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR U
; FILE REFERENCE: 01113,000IU3
; CURRENT APPLICATION NUMBER: US/09/991,258
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 09/902,537
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,995
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 1026
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
; NAME/KEY: CDS
; LOCATION: (1)...(1026)
US-09-991-258-9

Alignment Scores:
Pred. No.: 1,14 Length: 1026
Score: 74.50 Matches: 40
Percent Similarity: 37.42% Conservative: 21
Best Local Similarity: 24.54% Mismatches: 57
Query Match: 7.43% Indels: 45
DB: 10 Gaps: 11

US-10-008-355-2_COPY_522_712 (1-191) x US-09-991-258-9 (1-1026)
OY 38 ProGlyArg-----AlaLeuProSerAspAlaAsnPheThrMetArg 51
    |||||
Db 334 CCAGCGAAGACAGACGCGATGTCATGAATGGAATTCGACAAAGACGTTCCCAATC--- 390
OY 52 MetSerTrpGlySerLleLysGlyTrpGluProGlnAspGlyAla-----TrpTrpAsn 69
    |||||
Db 391 ATCTTGAAGGGAAGATTAACGCTACGCTTGCTGTGCTGGAGGGAAGTATTCAGAGCGG 450
OY 70 TyrHisThrThrGlyLys-----GlyValLeuGluLysGlnAsp 82
    |||||
Db 451 ATGCAATGTGAAGGCAAGATCGACACAGACGCTTCTGCGCGCTTAAAGCAAGAAAGCA 510
```

Oy	83	ProlysSerAsp-----GluPheAla---ValGIInGLuInsIle----LeuAspleuPheArg	99
Dd	511	TCCAAATTCAGTCTTTGGGTATGCAGATGCCAAGAACAATGGCGGGCCATTACATTCCAA	570
Oy	100	-----ThrllysAsnTYrGlYArGYtALa----	Glu 108
Dd	571	TACACCACATGAANAACCCCAAGGCTATTATCACGCTGCATCATGGACAGCAGTCCAATTAGA	630
Oy	109	AsnGLylglnLeuHISlleAlaPheLeuSerSernaSnaspIlleThrgllyGlyasnsergly	128
Dd	631	AATGGCGCTTCACAGTG-----CCGAAGAGACTTGGGGCCAGGGAGAACACGCCGA	681
Oy	129	SerProvalPheaasplysAsnGLyArgleuIIeclyleuAlaPheasrgly-----	145
Dd	682	CGACCATTCTTGATAACCAAGCAGCGGGTGCTGCTATTGGTGGGAGSGTGTGAATGA	741
Oy	146	-----AsnTrIpeluaImetSerGlyAspIIeGluPhe	156
Dd	742	GGATCTTAGACAGCCCTTTCAgtcgcTcatGTggNAACAgAAGGagAttTAaccttgAAgAtat	801
Oy	157	GluProasp	159
Dd	802	ACTCGGAG 810	

```

RESULT 13
US-09-991-258-8
: Sequence 8, Application US/09991258
: Patent No. US20020141975A1
: GENERAL INFORMATION:
: APPLICANT: Olmsted, Robert
: APPLICANT: Keith, Paula
: APPLICANT: Dryga, Sergey
: APPLICANT: Caley, Ian
: APPLICANT: Maughan, Maureen
: APPLICANT: Johnston, Robert
: APPLICANT: Davis, Nancy
: APPLICANT: Swanstrom, Ronald
: TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE
: FILE REFERENCE: 01113.0001U3
: CURRENT APPLICATION NUMBER: US/09/991,258
: CURRENT FILING DATE: 2001-11-16
: PRIOR APPLICATION NUMBER: 09/902,537
: PRIOR FILING DATE: 2001-07-09
: PRIOR APPLICATION NUMBER: 60/216,995
: PRIOR FILING DATE: 2000-07-07
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8
: LENGTH: 5076
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1 =
US-09-991-258-8

```

Alignment Scores:	
Pred. No.:	10.9
Score:	74.50
Percent Similarity:	37.42%
Best Local Similarity:	24.54%
Query Match:	7.43%
DB:	10
Gaps:	11
Length:	5076
Matches:	40
Conservative:	21
Mismatches:	57
Indels:	45
Gaps:	11

US-10-008-355-2_COPY_522_712 (1-191) x US-09-991-258-8 (1-5076)

0y 38 ProglyArg-----AlaLeuProSerAspAlaAsnPhetHmetArg 51
 |||||::: ||| ||||| ||| :::
 Db 909 CCAGGCAAGAGACACAGCGATGGTCATGAATTGGAACTGCACAGACGCTCCCAATC-- 965
 0y 52 MetSerTYRglySerIleIySGLTYRGIubProGlnAspGlyAla-----TrrPYrAsn 69

Db	966	ATGTGGCAAGGGAAGATAAACGGCTACGCTTGtGTGTGGAGGGAAGTATTATTCAGCGCG	1025
Qy	70	TyrHisThrThreGlyLys-----GlyValLeuGluLysGlnAsp	82
Db	1026	ATGCATGTGGAGGCAAGATCGACACAGAGCTTCGTGGCGCGCTTAAGACGAGAAGACGA	1085
Qy	83	ProLysSerAsp---GluPheAla---ValGlnGluAsnIle---LeuAspLeuPheArg	99
Db	1086	TCCAAATACGACTCTTGATGTGCAGATGTGGCCACAGAACATCGGGCCGATACATTCCAA	1145
Qy	100	-----ThrLysAsnTyrGluAArgTyrAla-----Glu	108
Db	1146	TACACCCATGAGAAACCCCAAGGCTATTATCACGTGGCATCATGACGACGTCCATATGAA	1205
Qy	109	AsnGlyGlnLeuHisIleAlaPheLeuSerAsnAsnAspIleThrGlyGlyAsnSerGly	128
Db	1206	AATGGCGCTTTCACCGTG-----CCGAAAGAGTTGGGGCCCAAGGAGACACGGCGA	1256
Qy	129	SerProValPheAspLysAsnGlyArgLeuIleGlyLeuAlaPheAspGly-----	145
Db	1257	CGACCCATTCTGGATTACCAAGGGACGGGTGTCGCTATTGTCTGGAGAGTGTGATGAA	1316
Qy	146	-----AsnTyrGluAlaIleMetSerGlyAspIleGluPhe	156
Db	1317	GGATCTAGAGACAGCCCTTTCATGTCGTCATGTGGAACGAGAAAGGAGTACCGTGAAGATAT	1376
Qy	157	GluProAsp	159
Db	1377	ACTCGCGAG	1385

```

RESULT 14
US-09-952-360-1/c
; Sequence 1, Application US/09952360
; Patent NO. US20020072066A1
GENERAL INFORMATION:
APPLICANT: Styrkarsdottir, Unnur
APPLICANT: Johannsdottir, Vala Drofn
TITLE OF INVENTION: Human Osteoporosis Gene
FILE REFERENCE: 2345, 2009-001
CURRENT APPLICATION NUMBER: US/09/952,360
CURRENT FILING DATE: 2001-09-13
PRIOR APPLICATION NUMBER: 09/661,867
PRIOR FILING DATE: 2000-09-14
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 14759
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (3639)...(3984)
NAME/KEY: CDS
LOCATION: (11757)...(12601)
US-09-952-360-1

```

Alignment Scores:	
pred. No.:	49.6
Score:	74.50
Percent Similarity:	37.30%
Best Local Similarity:	26.49%
Query Match:	7.43%
DB:	10
Gaps:	7
Length:	14759
Matches:	49
Conservative:	20
Mismatches:	70
Indels:	47

US-10-008-355-2_COPY_522_712 (1-191) x US-09-952-360-1 (1-14759)

OY 6 AAlaAalaaAglaAlaIeglnAlaapAlaMetAlaAsnAlaTyAlaIleGlu----- 23
:::||||| ::||| ||| |||
Db 1259 AGCGAGACGGGGCGAGCGCGCGGAGAGACTCCGCGCGGCAGCGCGCGCGC 1200
:::||||| ::||| ||| |||
OY 24 -----LysGlyLysArgLeuPheAlaGlyLeuArgGluMetCyrProGlyArg 40
:::||||| ||| ||| ||| ||| |||

Db 1199 GCGCCGTGGCGCGCTCGCGCTTTTAAAGGGAGCGCCGCTCCGCTCCCTCCGCG 1140
Qy 41 AlaleuProserAspAlaAsnPhetArgetArgetSerTyGlySerIleuSgLyTy 60
Db 1139 CCGCGCGCAACACCTCCCTCCCTCCGCGCGCTCCGCGCGCGCGCGCGCGCGCGCG 1081
Qy 61 GluProGlnAspGlyAlaArpTyraSerTyHisThrGlyLysGlyValLeuGluLys 80
Db 1080 GAGCTGCGCGGAGGTGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1021
Qy 81 GlnAspProLysSerAspGluPheAlaValGlnGluAsnIleuAspLeuPheArg 100
Db 1020 GCGGGGCGCGCGC----- 1009
Qy 101 LysAsnTyGlyArgTyraGlnAsnGlnLeuHisIleAlaPheLeuSerAsn 120
Db 1008 -----GCGCGCGCAAGGGGCGCGCTCCGCTCCAC---TCCCTGCTTCAAAAGCG 961
Qy 121 AspIleThrGly-----GlyAsnSerGlySerProValPheAspLys-As 135
Db 960 AGATGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 901
Qy 135 nGlyArgLeuIleGlyLeuAlaPheAspGlyAsnTrpGluAlaMetSer----- 151
Db 900 AGGGTCAAGGCTGCTGCTTATTTCC---CGTGGCGCGCTTATTCACCTTTCAGTT 844
Qy 152 -----GlyAspIleGluPheGlu-----ProAspLeuGlu 161
Db 843 GAAGTTTTCGTAATATACGATTTCACAAACAAACAAACAAACAAACCTCAATTACA 784
Qy 161 nArgThrIleSer 165
Db 783 GCGTATTTGGCA 771

RESULT 15

US-09-974-300-1065
; Sequence 1065, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085, 500-US
; CURRENT APPLICATION NUMBER: US/09/974, 300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680, 598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279, 526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8431
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1065
; LENGTH: 1472
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1065

Alignment Scores:

Pred. No.:	2.56	Length:	1472
Score:	73.50	Matches:	35
Percent Similarity:	38.41%	Conservative:	18
Best Local Similarity:	25.56%	Mismatches:	66
Query Match:	7.33%	Indels:	19
DB:	10	Gaps:	4

US-10-008-355-2_COPY_522_712 (1-191) x US-09-974-300-1065 (1-1472)

Qy 23 GluLysGlyLysArgLeu-----PhePheAlaGlyLeuArgGluMetTyProGly 39
Db 199 GAAATTGGCAACCGCTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 258
Qy 40 ArgAlaLeuProserAspAlaAsnPhetArgetArgetSerTyGlySerIleuSgLy 59

Db 259 TCTGCGCTGACACCGGAGCGCAATTTTACATG-----TTCATCCCGCGAG 303
Qy 60 TyrGluProGlnAspGlyAlaArpTyraSerTyHisThrGlyLysGlyValLeuGlu 79
Db 304 TACGAAACTGAGAGAGGTGCTTGAAGACGCGAGAGATCTCGGTGCTGCACTCCCTGC 363
Qy 80 LysGlnAspProLysSerAspGluPheAlaValGlnGluAsnIleuAspLeuPheArg 99
Db 364 GGTCA-----GAGTTCGCTGTCGCGAGCAT----- 393
Qy 100 ThrLysAsnTyGlyArgTyraGlnAsnGlnLeuHisIleAlaPheLeuSerAsn 119
Db 394 ---GAAAGACCGCGCGAGCTTGTGTCAGGGTCCGAGCGCGCTGTCTTACTAACCGCT 450
Qy 120 AsnAspIleThrGlyGlyLysAsnSerGlySerProValPheAspLysAsnGlyArgLeu 139
Db 451 TCTATATGTCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 510
Qy 140 GlyLeuAlaPheAspGlyAsnTrpGluAlaMetSerGlyAspIleGluPheGlu 157
Db 511 TTTCGCGAAGTGAATGCAATATACCGCAATGCTTGTGATGTGCTGAA 564

Search completed: December 21, 2002, 11:21:05
Job time : 71 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 13:23:05 : Search time 71 Seconds
(without alignments)
358.463 Million cell updates/sec

Title: US-10-008-355-2_COPY_522_712

Perfect score: 1003

Sequence: 1 SKSVTAARAIQADAMANAY.....LFMIDKGCPCRLIGELKLI 191

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_101002:*
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1003	100.0	712	23	AAO15205
2	274	27.3	52	23	AAO15206
3	90	9.0	711	17	AA88649
4	89.5	8.9	756	22	AA895249
5	89.5	8.9	856	22	AA894310
6	89.5	8.9	866	22	AA840262
7	89.5	8.9	923	22	AA842048
8	89.5	8.9	3421	22	AA842048
9	86	8.6	1060	23	AA853805
10	85.5	8.5	278	21	AA840033

11	85.5	8.5	321	21	AA840032
12	85.5	8.5	323	21	AA840031
13	84	8.4	537	18	AAW22220
14	83	8.3	201	22	AA82591
15	83	8.3	360	21	AA839512
16	82	8.2	870	21	AA841742
17	81.5	8.1	240	22	AAU33855
18	81.5	8.1	254	22	AAU36684
19	81.5	8.1	309	23	ABP25673
20	81	8.1	283	21	AA771478
21	81	8.1	283	23	AAU96101
22	80.5	8.0	532	17	AA891035
23	80.5	8.0	532	18	AAW22219
24	80.5	8.0	1317	22	ABG18723
25	80.5	8.0	4643	22	ABB71609
26	80	8.0	336	13	AA823644
27	80	8.0	344	17	AA891033
28	80	8.0	344	18	AAW22218
29	80	8.0	357	13	AA826842
30	80	8.0	392	17	AA891034
31	79.5	7.9	389	23	ABB53562
32	79	7.9	571	18	AAW63399
33	78.5	7.8	498	21	AAV75500
34	78	7.8	52	23	AAO15207
35	78	7.8	215	17	AA891044
36	78	7.8	509	22	AAW25688
37	78	7.8	684	22	AAW93433
38	78	7.8	781	21	AA831213
39	78	7.8	803	21	AA831212
40	78	7.8	846	21	AA831211
41	78	7.8	870	22	AA894408
42	78	7.8	3432	21	ABB07037
43	77.5	7.7	971	19	AAW48896
44	77	7.7	180	23	AAU75485
45	77	7.7	332	21	AAV68962

ALIGNMENTS

RESULT 1
AAO15205
ID AAO15205 standard; Protein; 712 AA.
AC AAO15205;
XX
DT 05-SEP-2002 (first entry)
XX
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7).
XX
KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KW DPP-7 inhibitor identification; periodontal disease; gingivitis;
XX
OS Porphyromonas gingivalis.
XX
PN WO200238742-A2.
XX
PD 16-MAY-2002.
XX
PF 08-NOV-2001; 2001WO-US46782.
XX
PR 08-NOV-2000; 2000US-246827P.
XX
(UYGE-) UNIV GEORGIA RES FOUND INC.
PA Travis J, Potempa JS, Banbula A, Bugno M;
PI WPI: 2002-490075/52.
XX N-PSDB: AAL43635.
DR Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT the dipeptidylpeptidase for protecting an animal from periodontal

PT disease caused by Porphyromonas gingivalis -
XX
PS Claim 7; Fig 4; 65pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present amino acid sequence represents the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme of the
CC invention.
XX
SQ Sequence 712 AA;

Query Match 100.0%; Score 1003; DB 23; Length 712;
Best Local Similarity 100.0%; Pred. No. 9.6e-101;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SRSVIAARAADAMANAAYAEKGRLEFFAGIREMYPGRALPSDANFTMRSSYSGIKGY 60
DB 522 SRSVIAARAADAMANAAYAEKGRLEFFAGIREMYPGRALPSDANFTMRSSYSGIKGY 581
QY 61 EPQDGMWYHRTGKGVLEKODPKSDEFAVOENITLDLFTKNYGRAENGOLHIAFLSN 120
DB 582 EPQDGMWYHRTGKGVLEKODPKSDEFAVOENITLDLFTKNYGRAENGOLHIAFLSN 641
QY 121 DITGNSGSPVFDKNGRLIGLAFDGNMEAMSGDIEFEPDLQRTISVDIRYVLFMIDKMQ 180
DB 642 DITGNSGSPVFDKNGRLIGLAFDGNMEAMSGDIEFEPDLQRTISVDIRYVLFMIDKMQ 701
QY 181 CPRLIOELKLI 191
DB 702 CPRLIOELKLI 712

RESULT 2
AA015206 standard; Protein: 52 AA.
XX
AC AA015206;
XX
DT 05-SEP-2002 (first entry)
XX
DE Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) C-terminal region.
XX
KW Dipeptidylpeptidase-7; DPP-7; enzyme; amidolytic cleavage;
KM DPP-7 inhibitor identification; periodontal disease; gingivitis;
KK periodontitis.
XX
OS Porphyromonas gingivalis.
XX
PN WO200238742-A2.
XX
PD 16-MAY-2002.
XX
PF 08-NOV-2001; 2001WO-US46782.
XX
PR 08-NOV-2000; 2000US-246827P.
XX
PA (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
PI Travis J, Potempa JS, Banbula A, Bugno M;
XX
DR WPI, 2002-490075/52.
XX
PT Novel isolated dipeptidylpeptidase useful for identifying inhibitor of
PT the dipeptidylpeptidase for protecting an animal from periodontal

PT disease caused by Porphyromonas gingivalis -
XX
PS Example 6; Fig 5; 65pp; English.
XX
CC The invention comprises the amino acid and coding sequence of the
CC Porphyromonas gingivalis dipeptidylpeptidase-7 (DPP-7) enzyme. The DPP-7
CC enzyme has amidolytic activity for cleavage of a peptide bond between the
CC second and third amino acids from the N-terminal end of a target peptide.
CC The DPP-7 target peptide has an aliphatic or aromatic residue as a
CC substituent on the alpha-carbon atom of the second amino acid from the
CC N-terminal end. The DPP-1 protein and DNA sequences of the invention are
CC useful for identifying a DPP-7 inhibitor. DPP-7 inhibitors are useful for
CC reducing the growth of a bacterium and protecting an animal from a
CC periodontal disease caused by Porphyromonas gingivalis (e.g. gingivitis
CC or periodontitis). The present amino acid sequence represents the
CC C-terminal region of the Porphyromonas gingivalis dipeptidylpeptidase-7
CC (DPP-7) enzyme.
XX
SQ Sequence 52 AA;

Query Match 27.3%; Score 274; DB 23; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.7e-22;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 TCGNSGSPVFDKNGRLIGLAFDGNMEAMSGDIEFEPDLQRTISVDIRYVLFM 174
DB 1 TCGNSGSPVFDKNGRLIGLAFDGNMEAMSGDIEFEPDLQRTISVDIRYVLFM 52

RESULT 3
AAR86649 standard; Protein: 711 AA.
XX
AC AAR86649;
XX
DT 02-SEP-1996 (first entry)
XX
DE Neisseria meningitidis BZ163 transferrin receptor Tbp2 subunit.
XX
KW Transferrin receptor; Tbp2 subunit; deletion mutant; vaccine;
KM passive immunisation; immunotherapy; IM2169; IM2394.
XX
OS Neisseria meningitidis (strain BZ163).
XX
FH Key
FH 1.711 Location/Qualifiers
FH Protein
FH /label= Tbp2-subunit
FH Peptide
FH 1..20
FH /label= signal_peptide
FH Domain
FH 21..366
FH /label= Domain_1
FH /note= "N-terminal domain"
FH 367..564
FH /label= Domain_2
FH /note= "hinge domain"
FH 565..711
FH /label= Domain_3
FH /note= "C-terminal domain"
XX
PN WO9533049-A2.
XX
PD 07-DEC-1995.
XX
PF 30-MAY-1995; 95WO-FR00701.
XX
PR 31-MAY-1994; 94FR-0006594.
XX
PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS.
XX (TRGE) TRANSCENE SA.
XX
PI Jacobs E, Legrain M, Mazarin V, Lissolo L;
PT Millet MBJ;

DR WPI; 1996-030562/03.
DR N-PSDB; AAT11244.
XX
PT Polypeptide(s) for vaccination against *Neisseria meningitidis* group
PT B - comprising deletion mutants of transferrin receptor Tbp2
PT subunit
XX
PS Disclosure; Page 82-87; 114pp; French.
XX
CC The present sequence is that of the *N. meningitidis* strain B2163
CC transferrin Tbp2 subunit. The Tbp2 polypeptide has three
CC domains (see features table); deletion mutants in which at least
CC one of the domains is partially or totally deleted are claimed,
CC provided that the first and second domains are not simultaneously
CC partially or totally deleted. The positions of the 3 domains in
CC B2163 are defined by alignment with the IM2169 sequence. The
CC deletion mutant polypeptides of the invention can generate an
CC immune response against *N. meningitidis*.
XX
SQ Sequence 711 AA;
XX
Query Match 9.0%; Score 90; DB 17; Length 711;
Best Local Similarity 21.8%; Pred. No. 1.2; Mismatches 66; Indels 42; Gaps 7;
Matches 38; Conservative 28;
XX
OY 2 KSVTAARAIQADAMANAIAIEKGRLEFAGLREMYPGRALPSDANFTMMSYGIKYE 61
DB 518 KNSKSMQAGESSQADAKTEQYGQSMFLQG--ERTDEKTEPSEQNIYV----- 565
XX
OY 62 PDGAMYNV---HTGKG-VLEKQDKPSDEFAVOENILDFRTKNYGRAENGOLHIAFL 117
DB 566 ---GSWYGHASSTSWSGNASDEKGNRAEFV-----NNGEKKITGL----- 606
XX
OY 118 SNNDINGGSGVFPDKNGRLGLAFDGNWEMSGDIEFEPDLQRTISVDIRV 171
DB 607 -----TAENRQETFTITDKIEGNGFSGT--AKTAELGFDLDKNTTRPKAVI 653
XX
RESULT 4
AAB95249
ID AAB95249 standard; Protein: 756 AA.
XX
AC AAB95249;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17413.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
PS Claim 8; SEQ ID 17413; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 756 AA;
XX
Query Match 8.9%; Score 89.5; DB 22; Length 756;
Best Local Similarity 25.1%; Pred. No. 1.5;
Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;
XX
OY 19 AYAIEKGRLEFAGLREMYPGRALPSDA--NFTMRM-----STGSIKGPDGAMYNV 71
DB 149 SYGVRGRGVCFEKINDEISVKHLPTSPDPHYVRIGWSLDSCTQIGEPFS---YGYG 205
XX
OY 72 TTGKGVLEKQDKPSDEFAVOENILDFRTKNYGRAENGOLHIAFLSNNDING----- 124
DB 206 GTGK-----KSTN-----SRFNYGDKFAE-----NDVIGCFADFE 236
XX
OY 125 -GNSGSPVFPDKNGRLGLAFDGNWEMSGDIEFEPDLQRTISVDIRV 170
DB 237 CGNDVELSTFTKNGKMWGIAIRIQKEALGQALPHYLKNCAYEFNF 283
XX
RESULT 5
AAB94310
ID AAB94310 standard; Protein: 856 AA.
XX
AC AAB94310;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:14779.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
XX
PR 02-MAY-2000; 2000JP-0183767.
XX
PR 09-JUN-2000; 2000JP-0241899.
XX

PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS
 XX Claim 8: SEQ ID 14779; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 856 AA:
 SQ
 Query Match 8.9%; Score 89.5; DB 22; Length 856;
 Best Local Similarity 25.1%; Pred. No. 1.8;
 Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;
 QY 19 AYAIEKGRLEFFAGLEMYPGRALPSDA--NFTMRM----SYGSIKGYEPDGDGAWYNYH 71
 DB 249 SYGVRGRVCFEEMKINEISVKHLPTSTEDDPHYVRIGWELDSCSTQLGEBEPS---YGVG 305
 QY 72 TTGKGYLEKQDPRKSDPEFANQENILDFRTKNGG-KRAENGQHLIAFLSNDDTG----- 124
 DB 306 GTGK-----KSTN-----SRENYGDKFAE-----NDVIQCFADPE 336
 QY 125 -GNSGSPVDFDKNGRLIGLAFDGNWEMASGDIEFEPDLORTISVDIRY 170
 DB 337 CGNDVELSTTKNGKMGIAFRIOKEALGQALYPHYLVKNCVERNF 383
 RESULT 6
 AAM40262
 ID AAM40262 standard: Protein; 866 AA.
 XX
 AC AAM40262;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 3407.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX

OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 XX 26-DEC-2000; 2000WO-US34263.
 PF
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0596042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HVSEQ INC.
 XX
 PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB: AA159418.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS
 XX Example 5: SEQ ID NO 3407; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/Inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 XX
 XX Sequence 866 AA:
 SQ
 Query Match 8.9%; Score 89.5; DB 22; Length 866;
 Best Local Similarity 25.1%; Pred. No. 1.8;
 Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;
 QY 19 AYAIEKGRLEFFAGLEMYPGRALPSDA--NFTMRM----SYGSIKGYEPDGDGAWYNYH 71
 DB 249 SYGVRGRVCFEEMKINEISVKHLPTSTEDDPHYVRIGWELDSCSTQLGEBEPS---YGVG 305
 QY 72 TTGKGYLEKQDPRKSDPEFANQENILDFRTKNGG-KRAENGQHLIAFLSNDDTG----- 124
 DB 306 GTGK-----KSTN-----SRENYGDKFAE-----NDVIQCFADPE 336
 QY 125 -GNSGSPVDFDKNGRLIGLAFDGNWEMASGDIEFEPDLORTISVDIRY 170
 DB 337 CGNDVELSTTKNGKMGIAFRIOKEALGQALYPHYLVKNCVERNF 383
 RESULT 7
 AAM42048
 ID AAM42048 standard: Protein; 923 AA.
 XX
 AC AAM42048;
 XX

DT 22-OCT-2001 (first entry)
XX Human polypeptide SEQ ID NO 6979.
DE
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YF, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AAI61204.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 6979; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 923 AA:
Query Match 8.9%; Score 89.5; DB 22; Length 923;
Best Local Similarity 25.1%; Pred. No. 2;
Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;

OY 125 -GNSGSPVEDKNGRLIGLAFDGNWEAMSGDIEEPDLORTISVDIR 170
DB 394 CGNDDELSTFKNGKMGKMGIAFRIOKEALGGQALYPHYLVKNCVAERNF 440
RESULT 8
AAB84902
ID AAB84902 standard; Protein; 3421 AA.
AC AAB84902;
XX
XX 23-JUL-2001 (first entry)
XX
XX Chimeric protein from yellow fever virus and Japanese encephalitis virus.
DE
XX
XX yellow fever virus; prM-E protein; flavivirus; chimeric; medicament;
KW infection; tumor antigen; cytokine; lymphoid; reticuloendothelial;
KW cancer; virucide; vaccine.
XX
XX Synthetic.
OS
OS Flavivirus febrilis.
OS Japanese encephalitis virus.
XX
PN WO200139802-A1.
XX
PD 07-JUN-2001.
XX
PF 01-DEC-2000; 2000WO-US32821.
XX
PR 01-DEC-1999; 990US-0452638.
XX
PA (ORAV-) ORAVAX INC.
XX
PI Chambers TJ, Monath TP, Guirakhoo F;
XX
DR WPI: 2001-343953/36.
DR N-PSDB: AAF83821.
XX
PT Chimeric live, infectious, attenuated yellow fever viruses used for
PT preventing and treating diseases caused by flaviviruses have prM-E
PT nucleotide sequence from a second, different flavivirus as functional
PT yellow fever prM-E is not expressed -
XX
PS Disclosure: Page 200-215; 232pp; English.
XX
XX The invention relates to a chimeric live, infectious, attenuated virus
CC comprising a yellow fever virus with the nucleotide sequence encoding a
CC prM-E protein deleted, truncated or mutated so that functional yellow
CC fever virus prM-E protein is not expressed and also integrated into the
CC genome of the yellow fever virus a nucleotide sequence encoding a prM-E
CC protein of a second, different flavivirus so that the prM-E protein of
CC the second flavivirus is expressed. The chimeric live, infectious,
CC attenuated virus is used to prepare medicaments for preventing or
CC treating flavivirus infection in a patient. The yellow fever virus vector
CC produces its gene product (tumor antigen or cytokine) in cells of the
CC lymphoid or reticuloendothelial system or in a precursor of these systems
CC in patients with cancer. Flaviviruses replicate in the cytoplasm of cells
CC so that the virus replication does not involve integration of the viral
CC genome into the host cell. The present sequence represents a chimeric
CC protein derived from yellow fever virus and Japanese encephalitis virus.
XX
SQ Sequence 3421 AA:
Query Match 8.9%; Score 89.5; DB 22; Length 3421;
Best Local Similarity 25.5%; Pred. No. 14;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

Db 1596 AVPGKNVNVQTKPS-----LFKVRNG---EIGAVAL-----DYPSTGSGSP 1635
Qy 131 VEDKNGRLGLAFDGNWEAMSGDIEFEPDLORT 163
Db 1636 IYNRNGEVLGLYNG---ILVGDNSFVSAISQT 1665

RESULT 9
ABBS3805
ID ABBS3805 standard; Protein: 1060 AA.
XX
XX ABB53805;
AC

DT 16-MAY-2002 (first entry)
XX

DE Lactococcus lactis protein dnaE.
XX

XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX

OS Lactococcus lactis IL1403.
XX

PN FR2807446-A1.
XX

PD 12-OCT-2001.
XX

PF 11-APR-2000; 2000FR-0004630.
XX

PR 11-APR-2000; 2000FR-0004630.
XX

PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
XX

PI Bolotline A, Sorokine A, Renault P, Ehrlich SD;
XX

DR WPI: 2002-043418/06.
XX

PT New nucleotide sequence useful in the identification of Lactococcus
XX

PS Lactis and related species -
XX

XX Claim 6; SEQ ID No 507; 2504pp; French.
XX

CC The present invention is related to a Lactococcus lactis nucleotide
CC sequence (ABBS3805) and related proteins (ABBS3800-ABBS3821). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify Lactococcus lactis or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-OCT-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX

XX Sequence 1060 AA;
SQ

Query Match 8.6%; Score 86; DB 23; Length 1060;
Best Local Similarity 21.7%; Pred. No. 6;

Matches 45; Conservative 31; Mismatches 73; Indels 58; Gaps 8;

Qy 14 DAMANAYALEKGRFLFPAAGRLREMYGRALPSPANT---MMSGSGIKGYPOGAWYN 69
Db 743 DALENGFETEK-----PSINLMKIGDFVKNKKIRLGLAHVQGISRDIAKNTIV 789

Qy 70 YHTTKGVLEKODPKSDFAVOENILDL-----FRTKNGRYAENQGLHIAFLS--NN 120
Db 790 ENQPKDLADFEVKEKLPNNFKKENILPLIQIGAFYADSNRGLAYNLADHANLINTYSD 849

Qy 121 DTGNGSGSPV-----FDKNGRLGLA-----FDGNWEAMSGD 153
Db 850 DIFMASGSGFAYHAEDYSETEKXDFEKN--LTGIGVTPHPLOGLARREFEENFPPLA-- 905

Qy 154 IEFEPDLORTISVDIRYVLFMDKMGQ 180
Db 906 -OLVKNRMRTILVEINLYIRTRTKTGQ 931

RESULT 10
AAG40033
ID AAG40033 standard; Protein: 278 AA.
XX

AC AAG40033;
XX

DT 18-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49616.
XX

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

OS Arabidopsis thaliana.
XX

PN EP1033405-A2.
XX

PD 06-SEP-2000.
XX

PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
XX

PR 05-MAR-1999; 99US-0123180.
XX

PR 09-MAR-1999; 99US-0123548.
XX

PR 23-MAR-1999; 99US-0125788.
XX

PR 25-MAR-1999; 99US-0126264.
XX

PR 29-MAR-1999; 99US-0126785.
XX

PR 01-APR-1999; 99US-0127462.
XX

PR 06-APR-1999; 99US-0128234.
XX

PR 08-APR-1999; 99US-0128714.
XX

PR 16-APR-1999; 99US-0129845.
XX

PR 19-APR-1999; 99US-0130077.
XX

PR 21-APR-1999; 99US-0130449.
XX

PR 23-APR-1999; 99US-0130510.
XX

PR 28-APR-1999; 99US-0130891.
XX

PR 30-APR-1999; 99US-0131449.
XX

PR 30-APR-1999; 99US-0132048.
XX

PR 04-MAY-1999; 99US-0132407.
XX

PR 05-MAY-1999; 99US-0132484.
XX

PR 06-MAY-1999; 99US-0132485.
XX

PR 06-MAY-1999; 99US-0132486.
XX

PR 07-MAY-1999; 99US-0132487.
XX

PR 11-MAY-1999; 99US-0132863.
XX

PR 14-MAY-1999; 99US-0134256.
XX

PR 14-MAY-1999; 99US-0134218.
XX

PR 14-MAY-1999; 99US-0134221.
XX

PR 14-MAY-1999; 99US-0134370.
XX

PR 18-MAY-1999; 99US-0134768.
XX

PR 19-MAY-1999; 99US-0134941.
XX

PR 20-MAY-1999; 99US-0135124.
XX

PR 21-MAY-1999; 99US-0135353.
XX

PR 24-MAY-1999; 99US-0135629.
XX

PR 25-MAY-1999; 99US-0136021.
XX

PR 27-MAY-1999; 99US-0136382.
XX

PR 28-MAY-1999; 99US-0136782.
XX

PR 01-JUN-1999; 99US-0137222.
XX

PR 03-JUN-1999; 99US-0137528.
XX

PR 04-JUN-1999; 99US-0137502.
XX

PR 07-JUN-1999; 99US-0137724.
XX

PR 08-JUN-1999; 99US-0138094.
XX

PR 10-JUN-1999; 99US-0138540.
XX

PR 10-JUN-1999; 99US-0138847.
XX

PR 14-JUN-1999; 99US-0139119.
XX

PR 16-JUN-1999; 99US-0139452.
XX

PR 16-JUN-1999; 99US-0139453.
XX

PR 17-JUN-1999; 99US-0139492.
XX

PR 18-JUN-1999; 99US-0139454.
XX

PR 18-JUN-1999; 99US-0139455.
XX

PR 18-JUN-1999; 99US-0139456.
XX

DB 243 KSGMSSGVNF-----AIPDVTWRTPEYLI 268

RESULT 11

AAc40032

AAc40032 standard; Protein: 321 AA.

XX

AC AAc40032;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49615.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EPI03405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 2000EP-0301439.

XX

PR 25-FEB-1999; 99US-0121825.
05-MAR-1999; 99US-0123180.
09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

```
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 8.5%; Score 85.5; DB 21; Length 321;

Best Local Similarity 27.6%; Pred. No. 1.2;

Matches 42; Conservative 16; Mismatches 67; Indels 27; Gaps 8;

```
OY 45 DANFTIRMSYGSJKYEP-ODGAWMYHTTGK-----GVLEKQDPKSDEFAY-----Q 91
DB 166 DANGTFRSREKGIYGLDPNDLAVLEKIEFEGRELNPVILGTSDNLRVGOSCFAGNPGY 225
OY 92 ENLIDFRTKNGRY--AENGQ-LHTAFLSNNDITGSGSPVFDKNGRLGL---AFDG 145
DB 226 ENLTITGVVSGCLRELPSPGKISIPAIOTDADINSGSGPLDSDGHTIGVNTATFTR 285
OY 146 NMEAMSGDIEFPDLORTISVD--IRYVLEMI 175
```

```
DB 286 KSGMSSGVNF-----AIPDITVVFRTVPYLI 311
RESULT 12
ID AAG40031 standard; Protein; 323 AA.
XX AAG40031;
AC AAG40031;
XX 18-OCT-2000 (first entry)
DT 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 49614.
DE Arabidopsis thaliana protein fragment SEQ ID NO: 49614.
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
XX EPI033405-A2.
XX 06-SEP-2000.
PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-01223180.
PR 09-MAR-1999; 99US-0122548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
```

PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0138817.
PR 22-JUN-1999; 990S-0138899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 02-JUL-1999; 990S-0142154.
PR 06-JUL-1999; 990S-0142055.
PR 08-JUL-1999; 990S-0142390.
PR 09-JUL-1999; 990S-0142803.
PR 12-JUL-1999; 990S-0142920.
PR 13-JUL-1999; 990S-0142977.
PR 14-JUL-1999; 990S-0143542.
PR 15-JUL-1999; 990S-0143624.
PR 16-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 21-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 26-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 06-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148311.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.

PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149729.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0152363.
PR 13-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160981.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161992.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

Query Match 8.5%; Score 85.5; DB 21; Length 323;
Best Local Similarity 27.6%; Pred. No. 1.2;
Matches 42; Conservative 16; Mismatches 67; Indels 27; Gaps 8;

OY 45 DANFTMRMSYSGIKYEP-QDGCAMYNHTTGK-----GVLEKODPKSDEFAY-----Q 91
Db 168 DAKGTRFSKECKIVGLDNDNLAVLKIETEGRELNPVVLGTSNDLVRVQSCFPAIGNPYCY 227
OY 92 ENLLDFTKKNYGRY--AENGO-LHIAFLSNNDITGNSGSPVPEKNGRLTGL---AFDG 145
Db 228 ENFLTIGVSGIGREIPSPGSKSISEAIQTDDADINSNGSGPLDLSYGHITIGVMTATFTR 287

Oy 146 NMEAMSGDIEFEPDQRTISVD--IRVYLEMI 175
Db 288 KGSMSGSGVNF-----AIPDITVVRTPYLI 313

RESULT 13

AAW22220
ID AAW22220 standard; Protein: 537 AA.

XX AAW22220;

DT 11-SEP-1997 (first entry)

DE Protein encoded by pV8f construct.

XX Mutant; Staphylococcus aureus; V8 protease; enzyme; denaturation; primer;
KM truncation; wild type; PCR; polymerase chain reaction; amplification;
XX proteolytic activity; fusion protein; beta-galactosidase; urea.

OS Chimeric - Escherichia coli.
XX Chimeric - Staphylococcus aureus.

EH Key Location/Qualifiers

FT Region 1..100 /note= "E. coli beta-galactosidase"

FT Region 101..124 /note= "R6 linker sequence"

FT Region 125..339 /note= "truncated S. aureus V8 protease"

FT Region 342..365 /note= "R6 linker sequence"

FT Region 366..537 /note= "aminoglucoiside 3'-phosphotransferase"

XX EP745669-A2.

XX 04-DEC-1996.

PD 31-MAY-1996; 96EP-0303939.

PF 02-JUN-1995; 95JP-0170086.

XX (SUNR) SUNTORY LTD.

XX Obsuye K, Yabuta M;

PI WPI: 1997-013693/02.

XX Staphylococcus aureus V8 protease mutants - with increased
PT resistance to denaturation

PS Claim 8; Page 19-20; 42pp; English.

XX The invention relates to new mutant Staphylococcus aureus V8 proteases
CC which have enzyme activity even under environmental conditions which
CC promote protein denaturation. The mutants are based on 3 truncated V8
CC proteases lacking 48 (AAW22218), 56 (AAW22219) or 53 (AAW22220) amino
CC acids from the C-terminal of the wild type protease. The mutants also
CC contain amino acid substitutions, especially D44E, N71S and/or R147K.

CC The protein sequence shown here represents a chimeric protein
CC comprising a truncated Staphylococcus aureus V8 protease lacking the
CC prepro and C-terminal 53 amino acids linked, via a synthetic R6 linker,
CC downstream of the E. coli beta-galactosidase. Also included downstream of
CC the V8 protease fragment is a second R6 linker and a fragment of the

CC aminoglucoiside 3'-phosphotransferase protein. The chimeric sequence
CC was generated by restriction digestion and ligation from the V8PRT(-)
CC sequence (see AAW22218) by using a natural EcoRV site which removed a
CC further 8 amino acid from the C-terminus. This truncated V8 protease,
CC designated V8f, retains its level of activity in the presence of a
CC higher concentration of protein denaturant e.g. 5 M urea.

XX Sequence 537 AA;

Query Match 8.4%; Score 84; DB 18; Length 537;
Best Local Similarity 28.1%; Pred. No. 3.6;
Matches 18; Conservative 16; Mismatches 24; Indels 6; Gaps 1;
Oy 123 TGGNSGSPVFDKNGRLIGLAEDGNMEAMSGDIEFEPDQRTISVDIRVYLEMI-----D 176
Db 289 TGGNSGSPVFNENKNEVIGIHMGVFNENFNGAVFINENRNFKQIEDIHFEELRYRRHH 348
Oy 177 KMGQ 180
Db 349 RWGR 352
RESULT 14
AAG82591
ID AAG82591 standard; Protein: 201 AA.
AC AAG82591;
DT 03-SEP-2001 (first entry)
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2276.
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KM vaccination; endocarditis.
XX Staphylococcus epidermidis.
OS WO200134809-A2.
PD 17-MAY-2001.
PE 09-NOV-2000; 2000WO-US30782.
PR 09-NOV-1999; 99US-0164258.
PA (GLAX) GLAXO GROUP LTD.
PI Kimerly MJ;
XX WPI: 2001-316495/33.
DR N-PSDB: AAH53441.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX Claim 18; Page 606; 2188pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX Sequence 201 AA;

Query Match 8.3%; Score 83; DB 22; Length 201;
Best Local Similarity 24.5%; Pred. No. 1.1;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 17:19:13 ; Search time 21 seconds
(Without alignments)
152.468 Million cell updates/sec

Title: US-10-008-355-2_COPY_522_712

Perfect score: 1003
Sequence: 1 SKSVIAAARAIOADAMANA.....LFMDKMGCPRLIOELKLI 191

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 106657 seqs, 16763532 residues

Total number of hits satisfying chosen parameters: 106657

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCR_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCRUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1003	100.0	712	US-10-008-355-2	Sequence 2, Appl1
2	936	93.3	699	US-10-008-355-8	Sequence 8, Appl1
3	427.5	42.6	720	US-10-008-355-9	Sequence 9, Appl1
4	334	33.3	716	US-10-008-355-7	Sequence 7, Appl1
5	274	27.3	52	US-10-008-355-3	Sequence 3, Appl1
6	262.5	26.2	732	US-10-008-355-6	Sequence 6, Appl1
7	255.5	25.5	734	US-10-008-355-5	Sequence 5, Appl1
8	81.5	8.1	240	US-09-815-242-5351	Sequence 5351, Ap
9	81.5	8.1	254	US-09-815-242-12277	Sequence 12277, A
10	78	7.8	283	US-10-062-624-4	Sequence 4, Appl1
11	77	7.7	52	US-10-008-355-4	Sequence 4, Appl1
12	77	7.7	332	US-09-767-041-21	Sequence 21, Appl1
13	74.5	7.4	342	US-09-991-258-10	Sequence 10, Appl1
14	73	7.3	336	US-09-987-107-44	Sequence 44, Appl1
15	73	7.3	436	US-09-815-242-11154	Sequence 11154, A
16	73	7.3	1057	US-10-135-322-5	Sequence 5, Appl1
17	73	7.3	1057	US-10-135-322-24	Sequence 24, Appl1
18	73	7.3	1057	US-09-918-508-6	Sequence 6, Appl1
19	72.5	7.2	280	US-10-062-994-4	Sequence 4, Appl1

20	72.5	7.2	280	US-10-062-994-4	Sequence 4, Appl1
21	72.5	7.2	337	US-09-987-107-46	Sequence 46, Appl1
22	72	7.2	283	US-10-059-964-40	Sequence 40, Appl1
23	72	7.2	524	US-09-925-300-1538	Sequence 1538, Ap
24	71.5	7.1	289	US-10-135-322-10	Sequence 10, Appl1
25	70	7.0	567	US-09-815-242-13511	Sequence 13511, A
26	70	7.0	1356	US-09-757-781-2	Sequence 2, Appl1
27	69.5	6.9	341	US-09-815-242-11798	Sequence 11798, A
28	69	6.9	432	US-09-815-242-10274	Sequence 10274, A
29	68.5	6.8	1242	US-09-925-299-911	Sequence 911, App
30	68	6.8	428	US-09-829-0948-2	Sequence 2, Appl1
31	68	6.8	434	US-09-971-536-62	Sequence 62, Appl1
32	68	6.8	607	US-09-344-882-16	Sequence 16, Appl1
33	68	6.8	719	US-10-095-139-8	Sequence 8, Appl1
34	67.5	6.7	265	US-09-815-242-5859	Sequence 5859, Ap
35	67.5	6.7	265	US-09-815-242-12884	Sequence 12884, A
36	67.5	6.7	358	US-09-925-299-923	Sequence 923, App
37	67	6.7	276	US-10-062-994-18	Sequence 18, Appl1
38	67	6.7	276	US-09-846-808-19	Sequence 19, Appl1
39	67	6.7	281	US-10-062-624-9	Sequence 9, Appl1
40	67	6.7	281	US-10-059-964-2	Sequence 2, Appl1
41	67	6.7	281	US-10-059-964-2	Sequence 2, Appl1
42	67	6.7	432	US-09-815-242-13866	Sequence 13866, A
43	67	6.7	597	US-09-815-242-11617	Sequence 11617, A
44	66.5	6.6	871	US-09-886-468-21	Sequence 21, Appl1
45	66	6.6	359	US-09-925-297-718	Sequence 718, App

ALIGNMENTS

RESULT 1
US-10-008-355-2
Sequence 2, Application US/10008355
Patent No. US20020164759A1
GENERAL INFORMATION:
APPLICANT: Travis, James
APPLICANT: Potemba, Jan S
APPLICANT: Babulka, Agnieszka
TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
FILE REFERENCE: 235,00440101
CURRENT APPLICATION NUMBER: US/10/008,355
PRIOR FILING DATE: 2001-11-08
PRIORITY APPLICATION NUMBER: US 60/246,827
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2
LENGTH: 712
TYPE: PRT
ORGANISM: Porphyromonas gingivalis
US-10-008-355-2

Query Match 100.0%; Score 1003; DB 9; Length 712;

Best Local Similarity 100.0%; Pred. No. 3.7e-96;
Matches 191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	SKSVIAAARAIOADAMANAIAIEKGRLEFFAGLRKREYEGRALPSDANFTMRMSYSIKCY	60
DB	522	SKSVIAAARAIOADAMANAIAIEKGRLEFFAGLRKREYEGRALPSDANFTMRMSYSIKCY	581
QY	61	EPDGDAMVYHTTGKGVLEKDPKSDPEFAVOENITLDLFTKNGRYAENGOLHIFLSNN	120
DB	582	EPDGDAMVYHTTGKGVLEKDPKSDPEFAVOENITLDLFTKNGRYAENGOLHIFLSNN	641
QY	121	DTTGSGSPVFDKNGRLIGLAFDGNWEMASGDIIEFPDLOQTISVDIRYVLFMDKMGQ	180
DB	642	DTTGSGSPVFDKNGRLIGLAFDGNWEMASGDIIEFPDLOQTISVDIRYVLFMDKMGQ	701
QY	181	CPRLIOELKLI 191	
DB	702	CPRLIOELKLI 712	

RESULT 2
US-10-008-355-8
; Sequence 8, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246.827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 8
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-8

Query Match 93.3%; Score 936; DB 9; Length 699;
Best Local Similarity 100.0%; Pred. No. 3.2e-89;
Matches 178; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SSVIAAARAATQADAMANYATEKGRLEFFAGLRMPGRALPSDANFTMRMSYSGIKY 60
DB 522 SSVIAAARAATQADAMANYATEKGRLEFFAGLRMPGRALPSDANFTMRMSYSGIKY 581
OY 61 EPQDAMVYHTTGKGVLEKQDPKSDFAVOENIILDFRTKNYGRYAENGQHLIAFLSN 120
DB 582 EPQDAMVYHTTGKGVLEKQDPKSDFAVOENIILDFRTKNYGRYAENGQHLIAFLSN 641
OY 121 DITGNSGSPVVDKNGRLIGLAFDGNWEMASGDIIEFPDLQRTISVDIRYVLMIDKW 178
DB 642 DITGNSGSPVVDKNGRLIGLAFDGNWEMASGDIIEFPDLQRTISVDIRYVLMIDKW 699

RESULT 3
US-10-008-355-9
; Sequence 9, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246.827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 9
; LENGTH: 720
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-9

Query Match 42.6%; Score 427.5; DB 9; Length 720;
Best Local Similarity 47.4%; Pred. No. 1.9e-36;
Matches 81; Conservative 34; Mismatches 55; Indels 1; Gaps 1;

OY 22 IEKGRLEFFAGLRMPGRALPSDANFTMRMSYSGIKYEPQDAMVYHTTGKGVLEKQ 81
DB 549 ILRAQRTYIAGLEMDGDDQPPDANLRLFTYGVQKGYSPRDVYVYGHQTLDGWEKE 608
OY 82 DPKSDFAVOENIILDFRTKNYGRYAE-NGQHLIAFLSNNDITGNSGSPVVDKNGRLIG 140
DB 609 DPKSDFAVOENIILDFRTKNYGRYAE-NGQHLIAFLSNNDITGNSGSPVVDKNGRLIG 668

OY 141 LAFDGNWEMASGDIIEFPDLQRTISVDIRYVLMIDKWGCCRLIOELKLI 191
DB 669 LNFDRNMEGVGGDIQYLDIYQKRSITIVDIRYVLMIDKWGGCRLIDEMANIV 719

RESULT 4
US-10-008-355-7
; Sequence 7, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246.827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Xylella fastidiosa
US-10-008-355-7

Query Match 33.3%; Score 334; DB 9; Length 716;
Best Local Similarity 40.0%; Pred. No. 9.5e-27;
Matches 74; Conservative 40; Mismatches 53; Indels 18; Gaps 5;

OY 8 ARAIADAMANYATEKGRLEFFAGLRMPGRALPSDANFTMRMSYSGIKYEPQDAM 67
DB 546 ARAIADAMANYATEKGRLEFFAGLRMPGRALPSDANFTMRMSYSGIKYEPQDAM 590
OY 68 YNYHTTGKGVLEKQDPKSDFAVOENIILDFRTKNYGRYAEN-NGQHLIAFLSNNDITG 125
DB 591 YNYHTTGKGVLEKQDPKSDFAVOENIILDFRTKNYGRYAEN-NGQHLIAFLSNNDITG 649
OY 126 NSGSPVVDKNGRLIGLAFDGNWEMASGDIIEFPDLQRTISVDIRYVLMIDKWGCCRLI 185
DB 650 NSGSPVVDKNGRLIGLAFDGNWEMASGDIIEFPDLQRTISVDIRYVLMIDKWGCCRLI 709
OY 186 OELKLI 190
DB 710 OELKLI 714

RESULT 5
US-10-008-355-3
; Sequence 3, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
; FILE REFERENCE: 235.00440101
; CURRENT APPLICATION NUMBER: US/10/008.355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246.827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis
US-10-008-355-3

Query Match 27.3%; Score 274; DB 9; Length 52;
Best Local Similarity 100.0%; Pred. No. 5.1e-22;

Matches	52;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	123	TCGNSGSPYFDKNGRLTGLAFDGNNEAMSGDIEEPDQLQRTISVDIRYLEM	174						
Dd	1	TCGNSGSPYFDKNGRLTGLAFDGNNEAMSGDIEEPDQLQRTISVDIRYLEM	52						

```

: RESULT 6
: US-10-008-355-6
: Sequence 6, Application US/10008355
: Patent No. US20020164759A1
: GENERAL INFORMATION:
: APPLICANT: Travis, James
: APPLICANT: Potempa, Jan S
: APPLICANT: Babula, Agnieszka
: TITLE OF INVENTION: Dipeptidylpeptidases And Methods of Use
: FILE REFERENCE: 235.00440101
: CURRENT APPLICATION NUMBER: US/10/008.355
: CURRENT FILING DATE: 2001-11-08
: PRIOR APPLICATION NUMBER: US 60/246,827
: PRIOR FILING DATE: 2000-11-08
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 6
: LENGTH: 732
: TYPE: PRT
: ORGANISM: Shewanella putrefaciens
: US-10-008-355-6

```

Query Match	26.2%	Score 262.5	DB 9	Length 732
Best Local Similarity	30.0%	Pred. No. 2.6e-19		
Matches 67	Conservative 37	Mismatches 86	Indels 33	Gaps 4

[illegible]

```

RESULT 7
US-10-008-355-5
: Sequence 5, Application US/10008355
: Patent No. US20020164759A1
: GENERAL INFORMATION:
: APPLICANT: Travis, James
: APPLICANT: Potempa, Jan S
: APPLICANT: Banbula, Agnieszka
: TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use
: FILE REFERENCE: 225.00440101
: CURRENT APPLICATION NUMBER: US/10/008,355
: CURRENT FILING DATE: 2001-11-08
: PRIOR APPLICATION NUMBER: US 60/246,827
: PRIOR FILING DATE: 2000-11-08
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 5
: LENGTH: 734
: TYPE: PRT
: ORGANISM: Shewanella putrefaciens
: US-10-008-355-5

```

Best Local Similarity 32.68; Pred. No. 1.4e-18;
Matches 57; Conservative 39; Mismatches 62; Indels 17; Gaps 3

```

QY 19 AVAIKGRGLFVFGGLREMYPGRALPSDANFTMRSSYSGISIKGYEDOGAMYNHTTGKGL 78
    ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db 574 AYIEQGGPVI-----ADNSSLRTVGHGKISPKDGLVAVPPTRLRGITV 619

```

OY 79 EKODPKSSEFAVOENILDFRTKRYGRAENG--QLHIFELSNDDITGNSGSSEVFDKN 136
 | | | | | |
Db 620 QK-DTIGIDPDAKQOOLEIKQOYGDEIFYMKSIDSVPANFLSTLDTGGNSGSTLNGRA 678

Oy 137 RLGLAFGEENWMSGDIEEPDRTISVDIRVLFMIDKWCQCPRLQLCKIT 191
 | :| | | | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :
Dd 679 ELVGLIFPGYVESIIGGMAFDNEINRSHYDSRMMLWNKYLLDHADNLAEMEIV 733

RESULT 8
US-09-815-242-5351

Patent No. US20020061569A1
GENERAL INFORMATION:

APPLICANT: Ohlsen, Karl L.
APPLICANT: zyskind, Judith W.

APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of
TITLE OF INVENTION: Identification of

```

; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815
; CURRENT FILING DATE: 2001-03-21

```

; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-25

; PRIOR APPLICATION NUMBER: 60/242,578
 ;
 ; PRIOR FILING DATE: 2000-10-23
 ;
 ; PRIOR APPLICATION NUMBER: 60/253 625
 ;

; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110

```
;; SOFTWARE: FastSEQ for Windows Version
;; SEQ ID NO 5351
; LENGTH: 240
```

```

;      TYPE: PRT
;      ORGANISM: Staphylococcus aureus
US-09-815-242-5351

```

Query Match	8.1%	Score 81.5;	DB 10;	Length 240;
Best Local Similarity	23.0%;	Pred. No. 0.36;		
Matches 34;	Conservative 22;	Mismatches 59;	Indels 33;	Gaps 6

Qy 37 YGGRALPSDANFTMRMSYGIKGYEPQDGAM-YNYHTTGKG----VLEKDPKSDFAV 90
||| ||| :
Db 108 YPKEDVASVIQVEERALIERGPKCKFENNDNVTPEFYAAGAKAGADRIVIKIETYPHPKKNKYVL 167

QY 91 QENILDLFRTK-----YGRYAENGQLHIAFLSNDDITGSGSEVDKNGRLIGLAF--- 143
1 : : : : :
Db 168 HESTGPVMSVEGSISIVSAHTS-----GNSGSPVLNSNNELIHFASD 212

```
QY      144 ---DGNWEAMSGDIEFEPLQRTISVDI 168
          | | | : | | : : | : : |
Db      213 VKNDNRRNAYG--VFYFTPEIKKFAIENI 238
```

```
RESULT 9
US-09-815-242-12277
; Sequence 12277, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12277
; LENGTH: 254
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12277

Query Match
Best Local Similarity 8.1%; Score 81.5; DB 10; Length 254;
Matches 34; Conservative 22; Mismatches 59; Indels 33; Gaps 6;

QY 37 YGGRALPSDANFTMRMSYSGISIKYEPDQGM-YNYHTTSGK-----VLEKOPKSDPEFV 90
DB 122 YGKEDEVSVIYQVEARIENGPGFNNDVTFPKYAAGAKAGDRIVIGYPPHYKKKYYL 181
QY 91 QENILDLFTKTN---YGRYAENGQLHIAFLSNNDITGNSGSPVFDKNGRLIGLAF--- 143
DB 182 HESTGPMVSVEGSSIIYSAHTES-----GNSGSPVLNSNNEILGIHFASD 226
QY 144 ---DGNWEMASGDIIEFEPDLQRTISVDI 168
DB 227 VKNDNRNAYG--VYFTPEIKKFLAENI 252

RESULT 10
US-10-062-624-4
; Sequence 4, Application US/10062624
; Patent No. US20020115840A1
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-Kilodalton Immunodominant Protein
; FILE REFERENCE: D6152CIP2/D1
; CURRENT APPLICATION NUMBER: US/10/062,624
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 09/660,587
; PRIOR FILING DATE: 2000-09-12
; NUMBER OF SEQ ID NOS: 46
```

```
; SEQ ID NO 4
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-5 protein
US-10-062-624-4

Query Match
Best Local Similarity 8.1%; Score 81; DB 12; Length 283;
Matches 43; Conservative 20; Mismatches 67; Indels 30; Gaps 7;

QY 23 EKGKRLPAGLPEMTPGRALPS---DANFTMR---MSTGSLK--GYEPDQGMVNYHTTG 74
DB 58 EKKKTTVYVYGLKENNAGDAISSQSDPDFTINYSFKYASNFLEFVAVIG-----YSIG 112
QY 75 KVLKLEKOPKSDPEFAVOENIIDLFRKNYGRAENGQLHIAFLSNNDITGNSGSP---- 130
DB 113 SPRIE-----VEKSYEAFDYKKNPGDNRYKNGATRYCALSHQDADDDMTSATDKF 161
QY 131 VEDKNGRLIGLAFDGN--WEAMSGDIEFEPDLQRTISVDI 168
DB 162 VYLINGLLNISFMTNICYETASKNINPLSPYICAGIGTDL 201

RESULT 11
US-10-008-355-4
; Sequence 4, Application US/10008355
; Patent No. US20020164759A1
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S
; APPLICANT: Banbula, Agnieszka
; TITLE OF INVENTION: Dipeptidyl peptidases And Methods Of Use
; FILE REFERENCE: 235,00440101
; CURRENT APPLICATION NUMBER: US/10/008,355
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: US 60/246,827
; PRIOR FILING DATE: 2000-11-08
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-008-355-4

Query Match
Best Local Similarity 7.8%; Score 78; DB 9; Length 52;
Matches 16; Conservative 13; Mismatches 22; Indels 0; Gaps 0;

QY 123 TCGNSGSPVFDKNGRLIGLAFDGNWEAMSGDIEFEPDLQRTISVDIRYVLF 173
DB 1 TCGNSGSPVFNENEKNEYIGIHMGVFNENFGAVFINENVNFLLKONIEDIHF 51

RESULT 12
US-09-767-041-21
; Sequence 21, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUTS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; PRIOR FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc-feature
; OTHER INFORMATION: CPS2J
US-09-767-041-21

Query Match
Best Local Similarity 25.4%; Score 77; DB 10; Length 332;
Matches 31; Conservative 21; Mismatches 36; Indels 34; Gaps 7;

OY 73 TKGYLEKODPKSD-----EFAVOENTLDFR-----TKNYGRAENGOLHIAFL 117
DB 30 TLELELLDDSSDSDSTLCLEFAEDGRIKIFRLPNGSVSNARNG-1KNTANTIMFV 88
OY 118 SNNDDITGNSGSPVF-----DKNGRLIGLAFDGNMEAMSGDIEFEPDLORTISVDIR 169
DB 89 DSDDIIVGDNIVESLYTCKLENDSDLSGGLLA-TFDGNTQ-----ESELCK-CQIDLE 138
OY 170 YV 171
DB 139 EI 140

RESULT 13
US-09-991-258-10
; Sequence 10, Application US/09991258
; Patent No. US20020141975A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Robert
; APPLICANT: Keith, Paula
; APPLICANT: Dryga, Sergey
; APPLICANT: Caley, Ian
; APPLICANT: Maughan, Maureen
; APPLICANT: Johnston, Robert
; APPLICANT: Davis, Nancy
; APPLICANT: Swanson, Ronald
; TITLE OF INVENTION: ALPHAVIRUS VECTORS AND VIROSOMES WITH MODIFIED HIV GENES FOR USE
; FILE REFERENCE: 01113.000103
; CURRENT APPLICATION NUMBER: US/09/991,258
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 09/902,537
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: 60/216,995
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; No. US20020141975A1e =
US-09-991-258-10

Query Match
Best Local Similarity 7.4%; Score 74.5; DB 10; Length 342;
Matches 40; Conservative 21; Mismatches 57; Indels 45; Gaps 11;

OY 38 PER-----ALPSDANFTKMSYSGIKYEPDGA--WNYHTTGK-----GVLEKOD 82
DB 112 PEKROMWVKLESDETFPL-MLEGRINGYACVYGKLFERPMHVEKIDNDVLAALKTKKA 170
OY 83 PKSD-EFA-VOENI-LDLFR-----TKNYGRYA-----ENGOLHIAFLSNNDDITGNSG 128
DB 171 SKYDLEKADVPOMNADRTKTYTHEKPOGTYSMHNGAVQYENKRFYV---PKGVGAKGDSG 227
OY 129 SPVFDKNGRLIGLAFDG-----NWEAMSGDIEFEPD 159
```

```

DB 228 RPLDNOGRVAIVLGVNFGSRTALSYVMNKEKVTYKTPDE 270

RESULT 14
US-09-987-107-44
; Sequence 44, Application US/09987107
; Patent No. US20020156007A1
; GENERAL INFORMATION:
; APPLICANT: GRAVERSEN, Jonas
; APPLICANT: MOESTRUP, Soren
; TITLE OF INVENTION: APOLIPOPROTEINS ANALOGUES
; FILE REFERENCE: GRAVERSENIA
; CURRENT APPLICATION NUMBER: US/09/987,107
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/264,022
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: DK PA2001 00057
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: DK PA2000 01682
; PRIOR FILING DATE: 2000-11-10
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 44
; LENGTH: 336
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: pT7 H6ubiIxf Apo A-1 plasmid
US-09-987-107-44

Query Match
Best Local Similarity 7.3%; Score 73; DB 9; Length 336;
Matches 42; Conservative 21; Mismatches 54; Indels 48; Gaps 9;

OY 26 KRLEFAGIREMYGRALPSDAN-----FTMRMSYSGIKY---EPDGAWNYHTTG 74
DB 51 QRLIFAG-KQLBDGRTL-SDVNIQKESTLHLVRLRGSIIEGSGDEPPQSPW----- 101
OY 75 KGVLEKODPKSDPEFAVOENTLDFRKNYGR-----YAENGOLHIAFLSNNDDITGCG 126
DB 102 -----DKVKDLATVYVYDVL-----KDSGRDYVSQFEGSALGKOLNKLNDMD----- 144
OY 127 SGSPVFDKNGRLIGLAFDGNMEAMSGDIEFEPDLORTISVDIRYV 171
DB 145 SVTSTFSKLEBOIGPYTQEFWDNLEKETE---GLRQEMSKDLEEV 186

RESULT 15
US-09-815-242-11154
; Sequence 11154, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Zylsen, Karl L.
; APPLICANT: Zylskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
```

```

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11154
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(436)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-11154

```

```

Query Match          7.3%; Score 73; DB 10; Length 436;
Best Local Similarity 26.2%; Pred. No. 5.9;
Matches 33; Conservative 18; Mismatches 47; Indels 28; Gaps 5;

QY 30 FAGLEMTPGRALPSDANFTMTMSYSGIKGYEPDGCAMTNYHTTGKGVLEKODPKSDEFA 89
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 31 FVGL-----AAAPSGA-----STGSREALFLRDGKSRF--LGKGYLKAVAAVNNNEIA 76
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 90 -----VOENILDLFRTKNYGRYAENGOLHTAFLSNNDITGNSGSPVFDKNG 136
    : : : : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 77 QAIYVKDATNQAIEIDQIMIDLGTENKSNFGANATLAYS-LNNAKAAASKGLPIAYIA 135
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 137 RLIGLA 142
    | | |
DB .136 ELNGTA 141

```

Search completed: December 21, 2002, 01:47:39
Job time : 24 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 17:10:54 : Search time 28 Seconds
(without alignments)
200.706 Million cell updates/sec

Title: US-10-008-355-2_COPY_522_712
Perfect score: 1003
Sequence: 1 SKSVIAARAIQADAMANAY.....LFMIDKMGCPRLIQELKLI 191

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCYTOS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84	8.4	537	1 US-08-657-192-15	Sequence 15, Appl
2	83	8.3	360	4 US-09-134-001C-4357	Sequence 4357, Ap
3	82	8.2	872	1 US-08-491-357-3	Sequence 3, Appl
4	82	8.2	872	3 US-08-968-633-3	Sequence 3, Appl
5	82	8.2	872	3 US-09-196-466-3	Sequence 3, Appl
6	82	8.2	872	5 PCR-US96-10823-3	Sequence 3, Appl
7	81	8.1	283	4 US-09-660-587-4	Sequence 3, Appl
8	81	8.1	283	4 US-09-261-358A-4	Sequence 4, Appl
9	80.5	8.0	532	1 US-08-657-192-9	Sequence 9, Appl
10	80.5	8.0	532	1 US-08-523-373-7	Sequence 7, Appl
11	80	8.0	344	1 US-08-657-192-3	Sequence 3, Appl
12	80	8.0	344	3 US-08-523-373-5	Sequence 5, Appl
13	80	8.0	392	3 US-08-523-373-6	Sequence 6, Appl
14	78	7.8	215	3 US-08-523-373-24	Sequence 24, Appl
15	75	7.5	213	3 US-08-523-373-22	Sequence 22, Appl
16	75	7.5	214	3 US-08-523-373-23	Sequence 23, Appl
17	74	7.4	793	3 US-08-433-522A-10	Sequence 10, Appl
18	74	7.4	793	3 US-09-135-166-10	Sequence 10, Appl
19	74	7.4	793	4 US-08-942-046-10	Sequence 10, Appl
20	73.5	7.3	977	4 US-09-206-942-53	Sequence 53, Appl
21	73.5	7.3	983	4 US-09-206-942-51	Sequence 51, Appl
22	72.5	7.2	280	4 US-08-733-230-4	Sequence 4, Appl
23	72.5	7.2	280	4 US-08-953-326-4	Sequence 4, Appl
24	72.5	7.2	663	1 US-08-765-081-7	Sequence 7, Appl
25	72.5	7.2	663	3 US-09-098-082-7	Sequence 7, Appl
26	72.5	7.2	691	1 US-08-064-174-2	Sequence 2, Appl
27	72.5	7.2	691	1 US-08-066-167-4	Sequence 4, Appl

28	72.5	7.2	691	2 US-08-449-733-2	Sequence 2, Appl
29	72.5	7.2	692	3 US-08-448-194-62	Sequence 62, Appl
30	72.5	7.2	692	4 US-08-867-921-62	Sequence 62, Appl
31	72.5	7.2	711	1 US-08-487-890A-98	Sequence 98, Appl
32	72.5	7.2	711	2 US-08-478-435-98	Sequence 98, Appl
33	72.5	7.2	711	2 US-08-337-483-98	Sequence 98, Appl
34	72.5	7.2	711	2 US-08-478-373-98	Sequence 98, Appl
35	72.5	7.2	711	3 US-08-474-671-98	Sequence 98, Appl
36	72.5	7.2	711	3 US-08-483-577A-98	Sequence 98, Appl
37	72.5	7.2	711	3 US-08-448-194-8	Sequence 8, Appl
38	72.5	7.2	711	3 US-08-897-438-98	Sequence 98, Appl
39	72.5	7.2	711	4 US-08-867-921-8	Sequence 8, Appl
40	72.5	7.2	711	4 US-08-637-654-98	Sequence 98, Appl
41	72.5	7.2	711	4 US-08-649-518-98	Sequence 98, Appl
42	72	7.2	792	3 US-08-433-522A-8	Sequence 8, Appl
43	72	7.2	792	3 US-09-135-166-8	Sequence 8, Appl
44	72	7.2	792	4 US-08-942-046-8	Sequence 8, Appl
45	71.5	7.1	369	1 US-08-230-047-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-657-192-15
; Sequence 15, Application US/08657192
; Patent No. 5747321
; GENERAL INFORMATION:
; APPLICANT: YABUTA, Masayuki
; APPLICANT: OHSUYE, Kazuhito
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; TITLE OF INVENTION: PROTEASES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: BURNS, DOANE, SWECKER & MATTHEIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/657,192
; FILING DATE: 03-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 7-170086
; FILING DATE: 02-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Neuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 001560-264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 537 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-657-192-15
Query Match 8.4%; Score 84; DB 1; Length 537;
Best Local Similarity 28.1%; Pred. No. 0.92;
Matches 18; Conservative 16; Mismatches 24; Indels 6; Gaps 1;
QY 123 TGSNSSPFEDKNKGRIGLAFDGNMEASGDLEFFEDLRTISVDIRYVLFMI-----D 176

Db 289 TCGNSGSPVFNKNEVIGIHMGVNEFGAVFINENAVNFKLQIEDIHFEFLRYRRHH 348
QY 177 KWGO 180
Db 349 RWGR 352

RESULT 2
US-09-134-001C-4357
; Sequence 4357, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: CTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4357
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4357

Query Match
Best Local Similarity 8.3%; Score 83; DB 4; Length 360;
Matches 34; Conservative 15; Mismatches 46; Indels 44; Gaps 6;

QY 57 IKGYEQDQAMVNYHTGKGVLEKODPKSDEFAVOENILDLRT-KNY-----GRYAEK 109
Db 20 MKGAPMKNKIKKPFGLKSLKEYDLGEEF---EGLDFAMTLKKYKQGTPTPRHLEG 76
QY 110 GOLP-----HIAPLSNNDITGNSGSPFDKNGRLGLAFDG 145
Db 77 KNIALLFKTSRTIRTAFTVASIDGAHPEFLGKNDIQGKKES--VEDTAVVLGRMFDG 134

QY 146 NWEAMSGDIEEPDLORT 164
Db 135 -----IEFRGFSQKTV 145

RESULT 3
US-08-491-357-3
; Sequence 3, Application US/08491357
; Patent No. 5716782
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, Joanne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,357

; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 872 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-491-357-3

Query Match
Best Local Similarity 8.2%; Score 82; DB 1; Length 872;
Matches 43; Conservative 18; Mismatches 60; Indels 56; Gaps 9;

QY 3 SVIAAARAIQADMANAIAIEGKRLFFAGLRKEMTPG-----RAL 42
Db 580 TLVACSRAPVEDAKQLASFLHGNASLFRRTKAPGPGEGSSSLHLNPTDKASSIQSRPL 639

QY 43 PSDANFTRMYSGLKQ-YEQDQAM---YNY-HTTGKGVLEKODPKSDEFAVOENILDL 97
Db 640 PSPPKFT---SODSPDGOENSEGGMEDYDVHLQGEKEEFK-----TQELLEK 687

QY 98 FTKNYGRYAENGQLHIAFLS-----NNDITGNSGSPVDKNGRLGL 141
Db 688 GNIVRGK---GQLELQQLKQFERLQDEVSRRPIDHLANMTPAPLVP--GRIGGL 738

RESULT 4
US-08-968-633-3
; Sequence 3, Application US/08968633
; Patent No. 6100384
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, Joanne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/968,633
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/491,357
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 872 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: No. 6100384 Relevant
;   TOPOLOGY: No. 6100384 Relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
; US-08-968-633-3

Query Match      8.2%; Score 82; DB 3; Length 872;
Best Local Similarity 24.3%; Pred. No. 3.1;
Matches 43; Conservative 18; Mismatches 60; Indels 56; Gaps 9;

QY 3 SVIAARAIOADMANAVYAEKGRLEFAGLREMYG-----RAL 42
DB 580 TLVACGRAVPEDAKQLASFLHGNASLLFRRTKAPGPGPGSSSLHNPDKASSIOSRPL 639
QY 43 PSDANFTMRMSGYSGIKG-YEPDGDAM---YNY-HTTGKGVLEKODPKSDFAVOENILDL 97
DB 640 PSPPKFT---SQDSPGQYENSGGMEDYDYVHLQKKEFEK-----TQKELLEK 687
QY 98 FRKXNGRYAENGQLIHAFLS-----NNDITGNSGSPVFDKNGRLIGL 141
DB 688 GNIVROGK---GQLELQQLKQFERLEQEVSRPIDHDLNMTPAQPLVP--GRTGGL 738

RESULT 5
US-09-196-466-3
; Sequence 3, Application US/09196466
; Patent No. 6124434
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/196,466
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/491,357
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 872 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
```

```

; ANTI-SENSE: NO
; US-09-196-466-3

Query Match      8.2%; Score 82; DB 3; Length 872;
Best Local Similarity 24.3%; Pred. No. 3.1;
Matches 43; Conservative 18; Mismatches 60; Indels 56; Gaps 9;

QY 3 SVIAARAIOADMANAVYAEKGRLEFAGLREMYG-----RAL 42
DB 580 TLVACGRAVPEDAKQLASFLHGNASLLFRRTKAPGPGPGSSSLHNPDKASSIOSRPL 639
QY 43 PSDANFTMRMSGYSGIKG-YEPDGDAM---YNY-HTTGKGVLEKODPKSDFAVOENILDL 97
DB 640 PSPPKFT---SQDSPGQYENSGGMEDYDYVHLQKKEFEK-----TQKELLEK 687
QY 98 FRKXNGRYAENGQLIHAFLS-----NNDITGNSGSPVFDKNGRLIGL 141
DB 688 GNIVROGK---GQLELQQLKQFERLEQEVSRPIDHDLNMTPAQPLVP--GRTGGL 738

RESULT 6
PCT-US96-10823-3
; Sequence 3, Application PC/TUS9610823
; GENERAL INFORMATION:
; APPLICANT: Golemis, Erica A.
; APPLICANT: Law, Susan
; APPLICANT: Estojak, JoAnne
; TITLE OF INVENTION: NUCLEIC ACID MOLECULE ENCODING A SIGNAL
; TITLE OF INVENTION: MEDIATOR PROTEIN THAT INDUCES CELLULAR MORPHOLOGICAL
; TITLE OF INVENTION: ALTERATIONS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
; STREET: 1601 Market Street Suite 720
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/10823
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Janet E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 872 amino acids
;   TYPE: amino acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: protein
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
; PCT-US96-10823-3

Query Match      8.2%; Score 82; DB 5; Length 872;
Best Local Similarity 24.3%; Pred. No. 3.1;
Matches 43; Conservative 18; Mismatches 60; Indels 56; Gaps 9;

QY 3 SVIAARAIOADMANAVYAEKGRLEFAGLREMYG-----RAL 42
DB 580 TLVACGRAVPEDAKQLASFLHGNASLLFRRTKAPGPGPGSSSLHNPDKASSIOSRPL 639
QY 43 PSDANFTMRMSGYSGIKG-YEPDGDAM---YNY-HTTGKGVLEKODPKSDFAVOENILDL 97
```

Db 640 PAPPKFT---SODSPDQGEENSEGGMEDDYVHLOGKEFEK-----: : : :
QY 98 FRTKNVGRANGOLHIAFLS-----NNDITGSGSPVFPKNGRLGL 141
Db 688 GNIVROGK---GQLELOQLKQFERLEQEVSRPIDHDLNMTPAQPLVP--GRTGGL 738

RESULT 7

US-09-660-587-4
; Sequence 4, Application US/09660587
; Patent No. 6392023
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP2
; CURRENT APPLICATION NUMBER: US/09/660,587
; PRIOR FILING DATE: 2000-09-12
; PRIOR APPLICATION NUMBER: 09/261,358
; PRIOR FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 46
; SEQ ID NO 4
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of p28-5 protein
US-09-660-587-4

Query Match 8.1%; Score 81; DB 4; Length 283;
Best Local Similarity 26.9%; Pred. No. 0.77; Matches 43; Conservative 20; Mismatches 67; Indels 30; Gaps 7;

QY 23 EKGKRLFFAGLEKREMPGRALPS---DANFTMR---MSYGSIK--GYEPDQGAANYHTTG 74
Db 58 EKKKTTVVYVGLKENNAGDAISSQSPDNDFTINYSFKYASNKFLGFAVAIG-----YSIG 112
QY 75 KGVLEKQDKPSDEFVAQENIIDLFRKKNYGRYAENGOLHIAFLSNNDITGNSGSP---- 130
Db 113 SPRIE-----VEMSYEAFVKNPGDNKNGAYRYCALSHQDDADDMTSATDKF 161
QY 131 VEDKNGRLGLAFDGN--WEAMSGDIEEPDLQRTISYDI 168
Db 162 VYLINEGLINISFMTNICYETASKNIPLSPIYICAGIGIDL 201

RESULT 8

US-09-261-358A-4
; Sequence 4, Application US/09261358A
; Patent No. 6403780
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: McBride, Jere W.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Homologous 28-kilodalton Immunodominant Protein
; TITLE OF INVENTION: Genes of Ehrlichia canis and Uses Thereof
; FILE REFERENCE: D6152CIP
; CURRENT APPLICATION NUMBER: US/09/261,358A
; PRIOR FILING DATE: 1999-03-03
; PRIOR APPLICATION NUMBER: 09/201,458
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 33
; SEQ ID NO 4
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Ehrlichia canis
; FEATURE:
; OTHER INFORMATION: amino acid sequence of Eca28SA2 protein
US-09-261-358A-4

Query Match 8.1%; Score 81; DB 4; Length 283;
Best Local Similarity 26.9%; Pred. No. 0.77; Matches 43; Conservative 20; Mismatches 67; Indels 30; Gaps 7;

QY 23 EKGKRLFFAGLEKREMPGRALPS---DANFTMR---MSYGSIK--GYEPDQGAANYHTTG 74
Db 58 EKKKTTVVYVGLKENNAGDAISSQSPDNDFTINYSFKYASNKFLGFAVAIG-----YSIG 112
QY 75 KGVLEKQDKPSDEFVAQENIIDLFRKKNYGRYAENGOLHIAFLSNNDITGNSGSP---- 130
Db 113 SPRIE-----VEMSYEAFVKNPGDNKNGAYRYCALSHQDDADDMTSATDKF 161
QY 131 VEDKNGRLGLAFDGN--WEAMSGDIEEPDLQRTISYDI 168
Db 162 VYLINEGLINISFMTNICYETASKNIPLSPIYICAGIGIDL 201

RESULT 9

US-08-657-192-9
; Sequence 9, Application US/08657192
; Patent No. 5747321
; GENERAL INFORMATION:
; APPLICANT: TABUTA, Masayuki
; APPLICANT: OHSUKE, Kazuhiko
; TITLE OF INVENTION: MUTANT STAPHYLOCOCCUS AUREUS V8
; TITLE OF INVENTION: PROTEASES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATTHIS
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/657,192
FILING DATE: 03-JUN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-170086
FILING DATE: 02-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-264
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO.: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-657-192-9

Query Match 8.0%; Score 80.5; DB 1; Length 532;
Best Local Similarity 30.5%; Pred. No. 2.2; Matches 18; Conservative 15; Mismatches 25; Indels 1; Gaps 1;

QY 123 TGGNSGSPVFDKNGRLGLAFDGNWEAMSGDIEEPDLQRTISYDI RYVLF-MIDKMGQ 180
Db 269 TCGNSGSPVFNKNEVIGIHMGVPRNFGAVFTINENVRNPLKQNIEDRLIRHHRWGR 347

RESULT 10

US-08-523-373-7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/523, 373
APPLICATION NUMBER: US/08/523, 373
FILING DATE: 05-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-238595
FILING DATE: 07-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 6-296028
FILING DATE: 07-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M.
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 001560-251
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6620
TELEFAX: 703-836-2021
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 213 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-523-373-22

Query Match 7.5%; Score 75; DB 3; Length 213;
Best Local Similarity 32.6%; Pred. No. 2.3;
Matches 15; Conservative 12; Mismatches 19; Indels 0; Gaps 0;
QY 123 TGGNGSGPVFDKNGRLIGLAFDGNMEAMSGDIEFEPPDLORTISVDI 168
|||||
DB 165 TGGNGSGPVFNEKNEVIGIHMGGVPPNEFGAVFINENVRNFLKONI 210
|||||

Search completed: December 21, 2002, 01:46:56
Job time : 31 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 15:49:10 ; Search time 45 Seconds

(without alignments)
408.037 Million cell updates/sec

Title: US-10-008-355-2_COPY_522_712

Perfect score: 1003

Sequence: 1 SKSVIAARAIQADAMANAY.....LFMDKMGCPRLIQELKLI 191

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334	33.3	716	2	G82627
2	95	9.5	507	2	T34152
3	93	9.3	513	2	D96980
4	90	9.0	711	2	S70660
5	89.5	8.9	856	2	T13159
6	89.5	8.9	3411	1	GNMVY
7	89.5	8.9	3411	1	GNMVY
8	86.5	8.6	576	2	B39228
9	86	8.6	1060	2	H86686
10	84	8.4	729	2	AE3591
11	84	8.4	968	2	S46992
12	83	8.3	698	2	D81832
13	82.5	8.2	759	2	AE1686
14	82	8.2	402	2	C87291
15	81.5	8.1	493	2	G90604
16	81.5	8.1	603	2	T04733
17	81	8.1	239	2	B89667
18	81	8.1	663	2	AC0782
19	81	8.1	1104	2	AE0999
20	80.5	8.0	325	2	E75404
21	80.5	8.0	434	2	A25525
22	80	8.0	336	1	PRASK
23	80	8.0	342	2	G89873
24	80	8.0	357	2	S21758
25	80	8.0	2043	2	T18524
26	80	8.0	3433	1	GNMVY
27	79.5	7.9	240	2	C89967
28	79.5	7.9	389	2	H86656
29	79.5	7.9	747	2	H82943

30	78.5	7.8	289	2	A89824	conserved hypothet
31	78.5	7.8	498	2	H81782	adhesin MaFB2 NMA2
32	78.5	7.8	608	2	T34994	probable long-chain
33	78.5	7.8	3388	1	GNMVY	genome polypeptide
34	78	7.8	271	2	T045780	hypothetical prote
35	78	7.8	846	2	C82931	hypothetical prote
36	78	7.8	1084	2	GNMVY	genome polypeptide
37	78	7.8	3432	1	GNMVY	hypothetical prote
38	77.5	7.7	515	2	A70905	hypothetical prote
39	77.5	7.7	647	1	G86118	2',3'-cyclic-nucle
40	77.5	7.7	647	1	G91277	2',3'-cyclic-nucle
41	77.5	7.7	3391	1	GNMVY	genome polypeptide
42	77	7.7	161	2	H84458	hypothetical prote
43	77	7.7	170	2	AD1934	hypothetical prote
44	77	7.7	332	2	C95077	pneumococcal surfa
45	77	7.7	412	2	H75484	hypothetical prote

ALIGNMENTS

```
RESULT 1
G82627
hypothetical protein XF1887 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82627
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: G82627
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-716 <SIM>
A:Cross-references: GB:AE004008; GB:AE003849; NID:g9106961; PIDN:AAF04693.1; GSPDB:GN
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Birones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.U.S.
Submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurane, E.E.; La
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tashiro, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
A:Genetics:
A:Gene: XF1887
Query Match 33.3%; Score 334; DB 2; Length 716;
Best Local Similarity 40.0%; Pred. No. 2.1e-22;
Matches 74; Conservative 40; Mismatches 53; Indels 18; Gaps 5;
```


Db 306 GCGK-----KSTN-----SREFNYGDKFAE-----NDVIGCFADFE 336
QY 125 -GNSGSPVFDKNGRLIGLAFDGNWEMSGDIEFEPDLQRTISVDIRY 170
Db 337 CGNDVELSTTKNGKMGIAFRIOKEFALGQALYPHVLRNCAVEFNF 383

RESULT 6
GNMVY

genome polypotein - yellow fever virus (strain 17D)
N:contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: yellow fever virus
C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 19-Jan-2001
C:Accession: A03914
R:Rice, C.M.; Lenches, E.M.; Eddy, S.R.; Shin, S.J.; Sheets, R.L.; Strauss, J.H.
Science 229, 726-733, 1985
A:Title: Nucleotide sequence of yellow fever virus: implications for flavivirus gene exp
A:Reference number: A03914; MUID:85272570; PMID:4023707
A:Accession: A03914
A:Molecule type: genomic RNA

A:Residues: 1-3411 <R1C>
A:Cross-references: GB:X03700; GB:K02749; NID:959338; PIDN:CAA27332.1; PID:959339
C:Superfamily: yellow fever virus genome polypotein
C:Keywords: ATP; capsid protein C #status predicted <CPC>
F:2-210/Product: capsid protein C #status predicted <CPC>
F:211-285/Product: envelope protein M #status predicted <EMP>
F:249-269/Domain: transmembrane #status predicted <TM2>
F:271-285/Domain: transmembrane #status predicted <TM1>
F:286-778/Product: major envelope protein E #status predicted <MEEP>
F:740-753/Domain: transmembrane #status predicted <TM3>
F:755-778/Product: transmembrane #status predicted <TM4>
F:779-1187/Product: nonstructural protein NS1 #status predicted <NS1>
F:1188-1354/Domain: transmembrane #status predicted <TM5>
F:1188-1354/Product: nonstructural protein NS2a #status predicted <N2A>
F:1355-1484/Product: nonstructural protein NS2b #status predicted <N2B>
F:1485-2107/Product: nonstructural protein NS3 #status predicted <NS3>
F:1682-1689/Region: nucleotide-binding motif A (P-loop)
F:1769-1774/Region: nucleotide-binding motif B
F:1773-1776/Region: DEAD motif
F:2108-2394/Product: nonstructural protein NS4a #status predicted <N4A>
F:2395-2506/Product: nonstructural protein NS4b #status predicted <N4B>
F:2507-3411/Product: nonstructural protein NS5 #status predicted <NS5>
F:134,150,172,266,594,755,908,986,1796,2062,2320,2346,2408,2467,2720,2734,2740/Binding site

Query Match 8.9%; Score 89.5; DB 1; Length 3411;
Best Local Similarity 25.5%; Pred. No. 31;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

QY 19 AVAIEGKRLFFAGLEMYGRALPSDANFTMRM-SYGSIKGYEPDQAMVNYH----- 71
Db 1542 AFLVRNGKRL-----IPSMASVKEDLVAYG---GSMWKEGRDGEDEVOLIA 1585
QY 72 -TTGKGVLEKQDKPSDEFAVOENILDLFRTKNYGRYAENGQHLIAFLNSNDITGNSGSP 130
Db 1586 AVPGKVVWVQTKPS-----LFKVRNGG---ELGAVL-----DYSGTSGSP 1625
QY 131 VFDKNGRLIGLAFDGNWEMSGDIEFEPDLQRT 163
Db 1626 IVNRNGEIVGLYNG---ILVGDNSFVSALSQT 1655

RESULT 7
GNMVY

genome polypotein - yellow fever virus (strain Pasteur 17D-204)
N:contains: capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5
C:Species: yellow fever virus
C:Date: 31-Mar-1991 #sequence_revision 31-Mar-1991 #text_change 19-Jan-2001
C:Accession: S07757
R:Dupuy, A.; Despres, P.; Cahour, A.; Girard, M.; Bouloy, M.
Nucleic Acids Res. 17, 3989, 1989

A:Title: Nucleotide sequence comparison of the genome of two 17D-204 yellow fever vac
A:Reference number: S07757; MUID:89282413; PMID:2734112
A:Accession: S07757
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: genomic RNA

A:Residues: 1-3411 <DDUP>
A:Cross-references: EMBL:X15062; NID:962289; PIDN:CA837419.1; PID:94456986
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1989, 1
C:Superfamily: yellow fever virus genome polypotein
C:Keywords: ATP; capsid protein C; envelope protein M; glycoprotein; nonstructural protei
F:2-210/Product: capsid protein C #status predicted <CPC>
F:105-125/Domain: transmembrane #status predicted <TM1>
F:211-285/Product: envelope protein M #status predicted <EMP>
F:271-289/Domain: transmembrane #status predicted <TM2>
F:286-778/Product: major envelope protein E #status predicted <MEEP>
F:736-753/Domain: transmembrane #status predicted <TM3>
F:756-778/Domain: transmembrane #status predicted <TM4>
F:779-1187/Product: nonstructural protein NS1 #status predicted <NS1>
F:1188-1354/Domain: transmembrane #status predicted <TM5>
F:1188-1354/Product: nonstructural protein NS2a #status predicted <N2A>
F:1355-1484/Product: nonstructural protein NS2b #status predicted <N2B>
F:1485-2107/Product: nonstructural protein NS3 #status predicted <NS3>
F:1682-1689/Region: nucleotide-binding motif A (P-loop)
F:1769-1774/Region: nucleotide-binding motif B
F:1773-1776/Region: DEAD motif
F:2108-2394/Product: nonstructural protein NS4a #status predicted <N4A>
F:2395-2506/Product: nonstructural protein NS4b #status predicted <N4B>
F:2507-3411/Product: nonstructural protein NS5 #status predicted <NS5>
F:134,150,172,266,594,908,986,1796,2062,2320,2346,2408,2467,2720,2734,2740/Binding site

Query Match 8.9%; Score 89.5; DB 1; Length 3411;
Best Local Similarity 25.5%; Pred. No. 31;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

QY 19 AVAIEGKRLFFAGLEMYGRALPSDANFTMRM-SYGSIKGYEPDQAMVNYH----- 71
Db 1542 AFLVRNGKRL-----IPSMASVKEDLVAYG---GSMWKEGRDGEDEVOLIA 1585
QY 72 -TTGKGVLEKQDKPSDEFAVOENILDLFRTKNYGRYAENGQHLIAFLNSNDITGNSGSP 130
Db 1586 AVPGKVVWVQTKPS-----LFKVRNGG---ELGAVL-----DYSGTSGSP 1625
QY 131 VFDKNGRLIGLAFDGNWEMSGDIEFEPDLQRT 163
Db 1626 IVNRNGEIVGLYNG---ILVGDNSFVSALSQT 1655

RESULT 8
B39228

flagellin B - Campylobacter jejuni
C:Species: Campylobacter jejuni
C:Date: 20-Sep-1991 #sequence_revision 20-Sep-1991 #text_change 26-Aug-1999
C:Accession: B39228
R:Nilfsten, P.J.M.; van Asten, F.J.A.M.; Gastra, W.; van der Zeijst, B.A.M.
J. Biol. Chem. 265, 17798-17804, 1990
A:Title: Structural and functional analysis of two Campylobacter jejuni flagellin gen
A:Reference number: A39228; MUID:91009243; PMID:2211662
A:Accession: B39228
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-576 <NU1>
A:Cross-references: GB:J05635; NID:9144197; PIDN:AAA23025.1; PID:9144199
C:Superfamily: flagellin

Query Match 8.6%; Score 86.5; DB 2; Length 576;
Best Local Similarity 31.5%; Pred. No. 5.9;
Matches 28; Conservative 15; Mismatches 39; Indels 7; Gaps 3;

QY 66 AMYNIHTTGKGVLEKQDKPSDEFAVOENILDLFRTKNYGRYAENGQHLIAFLNSNDITG 125
Db 232 ATYDVKTGTGYAL-KCCTTSDFAINGVVIGOT---NYKDDNGNQLVAINNAVKDTGTG 287

POLG_YEYV1 STANDARD: PRT: 3411 AA.
ID POLG_YEYV1 042028: Rel. 01. Created)
AC P0314: 21-JUL-1996 (Rel. 01. Last sequence update)
DT 21-JUL-1996 (Rel. 01. Last sequence update)
DT 15-JUN-2002 (Rel. 41. Last annotation update)
DE Genome polypeptide (Contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase
DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)
DE (NS5)).
OS Yellow fever virus (strain 17D).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85272570; PubMed=4023707;
RA Rice C.M., Lencches E.M., Eddy S.R., Shin S.J., Sheets R.L.,
RA Strauss J.H.;
RT "Nucleotide sequence of yellow fever virus: implications for
RT flavivirus gene expression and evolution.";
RL Science 229:726-733(1985).
CC - FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC - CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
CC precursor polypeptide, commonly with Asp or Glu in the p6
CC position, Cys or Thr in p1 and Ser or Ala in p1'.
CC - CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC {RNA}(N).
CC - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND MRA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).
CC -----
CC EMBL: X03700; CAA27332.1;
DR PIR: A03914; GNMVY.
DR HSSP: P14336; ISVB.
DR MEROPS: S07.001;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR000069; Flavi_M.
DR InterPro: IPR001157; Flavi_NS1.
DR InterPro: IPR000752; Flavi_NS2A.
DR InterPro: IPR000487; Flavi_NS2B.
DR InterPro: IPR000404; Flavi_NS4A.
DR InterPro: IPR001528; Flavi_NS4B.
DR InterPro: IPR000208; Flavi_NS5.
DR InterPro: IPR001122; Flavi_capsidC.
DR InterPro: IPR000336; Flavi_glycoprote.
DR InterPro: IPR001850; Flavi_helicase.
DR InterPro: IPR002553; Flavi_propep.
DR InterPro: IPR002877; FUSJ.
DR InterPro: IPR001650; Helicase_C.
DR Pfam: PF00271; helicase_C_1.
DR Pfam: PF00869; Flavi_glycoprot; 1.
DR Pfam: PF00948; Flavi_NS1; 1.
DR Pfam: PF00949; Flavi_helicase; 1.
DR Pfam: PF00972; Flavi_NS5; 1.
DR Pfam: PF01002; Flavi_NS2B; 1.
DR Pfam: PF01003; Flavi_capsid; 1.
DR Pfam: PF01004; Flavi_M; 1.
DR Pfam: PF01005; Flavi_NS2A; 1.
DR Pfam: PF01349; Flavi_NS4B; 1.
DR Pfam: PF01350; Flavi_NS4A; 1.

DR Pfam: PF01570; Flavi_propep; 1.
DR Pfam: PF01728; FUSJ; 1.
DR Pfam: PF02832; Flavi_glycop_C; 1.
DR ProDom: PD001496; Flavi_NS1; 1.
DR ProDom: PD001556; Flavi_glycoprote; 1.
DR SMART: SM00490; Helicase; 1.
DR KMPolyprotein; Glycoprotein; Transferrase; RNA-directed RNA polymerase;
KW Core protein; Coat protein; Envelope protein; Hydrolyase; Helicase;
KW ATP-binding; Transmembrane; Nonstructural protein.
FT INIT_MET 1
FT CHAIN 1
FT PROPEP 122 121
FT CHAIN 211 285
FT CHAIN 286 778
FT CHAIN 779 1130
FT CHAIN 1131 1354
FT CHAIN 1355 1484
FT CHAIN 1485 2107
FT CHAIN 2108 2256
FT CHAIN 2257 2506
FT CHAIN 2507 3411
FT TRANSMEM 249 269
FT TRANSMEM 271 285
FT TRANSMEM 740 753
FT TRANSMEM 755 778
FT TRANSMEM 1159 1180
FT DOMAIN 383 396
FT NP_BIND 1682 1689
FT SITE 1773 1776
FT DISULFID 288 315
FT DISULFID 345 401
FT DISULFID 359 390
FT DISULFID 377 406
FT DISULFID 467 568
FT DISULFID 585 615
FT CARBOHYD 134 134
FT CARBOHYD 150 150
FT CARBOHYD 908 908
FT CARBOHYD 986 986
FT CARBOHYD 2320 2320
FT CARBOHYD 2346 2346
FT CARBOHYD 2467 2467
SQ SEQUENCE 3411 AA; 379512 MW; 680E0EFCAD23DCFA6 CRC64;

Query Match 8.9%; Score 89.5; DB 1; Length 3411;
Best Local Similarity 25.5%; Pred. No. 10;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

QY 19 AVAIEGKRLFFAGLREMYGRALPSDANFTMRM-SYGSINKYEPQDGAANYNH----- 71
DB 1542 APLVNRGKKL-----IPSMASVKEDLVAVG---GSMKLEGRMDGEVEVQILTA 1585
QY 72 -TTGCGVLEKQPKSDPEFAVQENIIDLFRTKNYGRKAENGQHLIAFLNSNDITGNSGSP 130
DB 1586 AVPGKRVNVVQTRKPS-----LFLKVRNG---ELGAVAL-----DYPSTGSGSP 1625
QY 131 VPDKNRGLTGLAFDGNWEMASGDIEFEPDLOQT 163
DB 1626 IYNRNGEVITGLGNG---ILVVDNSFVSAISQT 1655

RESULT 3
POLG_YEYV2 STANDARD: PRT: 3411 AA.
ID POLG_YEYV2 01-FEB-1991 (Rel. 17. Created)
AC P19901: 01-FEB-1991 (Rel. 17. Last sequence update)
DT 01-FEB-1991 (Rel. 17. Last sequence update)
DT 15-JUN-2002 (Rel. 41. Last annotation update)
DE Genome polypeptide (Contains: Capsid protein C (Core protein); Matrix
protein (Envelope protein M); Major envelope protein E; Nonstructural
proteins NS1, NS2A, NS2B, NS4A and NS4B; Protease/helicase
DE (EC 3.4.21.98) (NS3); RNA-directed RNA polymerase (EC 2.7.7.48)

DE (NS5).
 OS Yellow fever virus (strain Pasteur 17D-204).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
 OX Flavivirus.
 RN NCBI_TaxID=11091;
 RP [1]
 RP SEQUENCE FROM N.A. PubMed=2734112;
 RA MEDLINE=89282413; PubMed=2734112;
 RA Dupuy A., Despres P., Gahour A., Girard M., Bouloy M.;
 RT "Nucleotide sequence comparison of the genome of two 17D-204 yellow
 RT fever vaccines";
 RT Nucleic Acids Res. 17:3989-3989(1989).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
 CC precursor polypeptide, commonly with Asp or Glu in the P6
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA}(N).
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPID-PROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X15062; CAB37419.1; .
 DR PIR: S07757; GNMWYP.
 DR HSSP: P14336; ISVP.
 DR MEROPS: S07.001; .
 DR InterPro: IPR001410; DPAAD.
 DR InterPro: IPR000069; Flavi_M.
 DR InterPro: IPR001157; Flavi_NSI.
 DR InterPro: IPR000752; Flavi_NSI.
 DR InterPro: IPR000487; Flavi_NS2B.
 DR InterPro: IPR000404; Flavi_NS4A.
 DR InterPro: IPR001528; Flavi_NS4B.
 DR InterPro: IPR000208; Flavi_NS5.
 DR InterPro: IPR001122; Flavi_capsid.
 DR InterPro: IPR000336; Flavi_glycoprote.
 DR InterPro: IPR001850; Flavi_glycoprote.
 DR InterPro: IPR002535; Flavi_helicase.
 DR InterPro: IPR002877; Flavi_helicase.
 DR InterPro: IPR001650; Flavi_helicase_C.
 DR Pfam: PF00271; helicase_C.1.
 DR Pfam: PF00869; Flavi_glycoprote.1.
 DR Pfam: PF00948; Flavi_NSI.1.
 DR Pfam: PF00949; Flavi_helicase.1.
 DR Pfam: PF00972; Flavi_NS5.1.
 DR Pfam: PF01002; Flavi_NS2B.1.
 DR Pfam: PF01003; Flavi_capsid.1.
 DR Pfam: PF01004; Flavi_M.1.
 DR Pfam: PF01005; Flavi_NS2A.1.
 DR Pfam: PF01349; Flavi_NS4B.1.
 DR Pfam: PF01350; Flavi_NS4A.1.
 DR Pfam: PF01570; Flavi_propep.1.
 DR Pfam: PF01728; Flavi_1.
 DR Pfam: PF02833; Flavi_glycop_C.1.
 DR Pfam: PF001496; Flavi_NSI.1.
 DR Pfam: PF001556; Flavi_glycoprote.1.
 DR SMART: SM00490; HELIC_C.1.
 KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase;
 KW Core protein; Coat protein; Envelope protein; Hydrolase; Helicase;
 KW ATP-binding; Transmembrane; Nonstructural protein.
 FT INIT_MET 1
 FT 1
 FT REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 121 CAPSID PROTEIN C.
 FT PROPEP 122 210 ENVELOPE GLYCOPROTEIN M.
 FT CHAIN 211 285 MAJOR ENVELOPE PROTEIN E.
 FT CHAIN 286 778 NONSTRUCTURAL PROTEIN NS1.
 FT CHAIN 779 1130 NONSTRUCTURAL PROTEIN NS2A.
 FT CHAIN 1131 1354 NONSTRUCTURAL PROTEIN NS2B.
 FT CHAIN 1355 1484 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 1485 2107 PROTEINASE/HELICASE (NS3).
 FT CHAIN 2108 2256 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2257 2506 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2507 3411 RNA-DIRECTED RNA POLYMERASE (NS5).
 FT NP_BIND 1682 1689 ATP (POTENTIAL).
 FT SITE 1773 1776 DEAH BOX.
 FT TRANSMEM 249 269 POTENTIAL.
 FT TRANSMEM 271 285 POTENTIAL.
 FT TRANSMEM 740 753 POTENTIAL.
 FT TRANSMEM 755 778 POTENTIAL.
 FT TRANSMEM 1159 1180 POTENTIAL.
 FT DISULFD 288 315 BY SIMILARITY.
 FT DISULFD 345 401 BY SIMILARITY.
 FT DISULFD 359 390 BY SIMILARITY.
 FT DISULFD 377 406 BY SIMILARITY.
 FT DISULFD 467 568 BY SIMILARITY.
 FT DISULFD 585 615 BY SIMILARITY.
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 150 150 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 908 908 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 986 986 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2320 2320 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2346 2346 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2467 2467 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 3411 AA; 379524 MW; 3298C0771FED23F7 CRC64;

Query Match 8.9%; Score 89.5; DB 1; Length 3411;
 Best Local Similarity 25.5%; Pred. No. 10;

Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

OY 19 AVAIEKGRLEFFAGLEMYPGHALPSDANFTWRM-SYSSIKYEPDQGAWYH----- 71
 DB 1542 AFLVANGKRL-----IPSWASVKEDLVAYG--GSMKLEGRMOGEERVOILA 1585
 OY 72 -TTGKGVLEKQDPKSDERAVIDENILDFRTKNYGYAENGQHLIAFLNSNDITGNSGSP 130
 DB 1586 AVPGKNVYVVCGRPS-----LEKVRNGG--EIGAVAL-----DYPSTGSSP 1625
 OY 131 VPDKNGRLIGLAFDGNWEMSGDIFEPDLORT 163
 DB 1626 IYNRNGEVIYGLYNGC---ILVGDNSFVSASISOT 1655
 RESULT 4
 FIB2_CAMUE STANDARD; PRT; 575 AA.
 AC P22252;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Flagellin B.
 GN FLAB.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=81116;
 RX MEDLINE=91009243; PubMed=2211662;
 RA Nuijten P.J., van Asten F.J., Gaastra W., van der Zeijst B.A.;
 RT "Structural and functional analysis of two Campylobacter jejuni
 RT flagellin genes";
 RL J. Biol. Chem. 265:17798-17804(1990).
 CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
 CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.

```
CC -1- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J05635; AAA23025.1; -.
CC PIR: B39228; B39228.
CC InterPro: IPR001492; FlagellinN.
CC InterPro: IPR001029; Flagellin_C.
CC Pfam: PF00669; Flagellin_N_1.
CC Pfam: PF00700; Flagellin_C_1.
CC PRINTS: PR00207; FLAGELLIN.
CC ProDom: PD000316; Flagellin_C_1.
CC Flagella.
CC KW FLAGELLIN.
CC FT INIT_MET 0 BY SIMILARITY.
CC SEQUENCE 575 AA; 59728 MW; D0531AF308A7BF1D CRC64;

Query Match 8.6%; Score 86.5; DB 1; Length 575;
Best Local Similarity 31.5%; Pred. No. 2.2;
Matches 28; Conservative 15; Mismatches 39; Indels 7; Gaps 3;

QY 66 AMYNHTGKGVLEKQDKSDFEAVQENILDFRTKNGRAENGQHLIAFLSNNDITGG 125
DB 231 AYDVKTTGVYAI-KEGTSQDFALNGVYIGQI---NKKDGNNGVLSAINAVKDTGV 286
QY 126 NSGSPFDKNGRLIGLAFDGNWEAMSGDI 154
DB 287 QASK---DENGKLVLTSDAGRGKIKITGDI 312

RESULT 5
DP3A_LACLA STANDARD; PRT; 1060 AA.
ID DP3A_LACLA
AC 09C170;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA polymerase III alpha subunit (EC 2.7.7.7).
GN DNAE OR IL0496.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Mincker P., Manger S., Tailon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
CC -1- FUNCTION: DNA POLYMERASE III IS A COMPLEX, MULTICHAIN ENZYME
CC RESPONSIBLE FOR MOST OF THE REPLICATIVE SYNTHESIS IN BACTERIA.
CC THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC THE ALPHA CHAIN IS THE DNA POLYMERASE (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -1- SUBUNIT: CONTAINS A CORE (COMPOSED OF ALPHA, EPSILON, AND THETA
CC CHAINS) THAT ASSOCIATES WITH A TAU SUBUNIT WHICH ALLOW THE CORE
CC DIMERIZATION TO FORM THE POLYIT COMPLEX. POLYT ASSOCIATES WITH
CC THE GAMMA COMPLEX (COMPOSED OF CHAINS GAMMA, DELTA, DELTA', PSI,
CC AND CHI) AND WITH THE BETA CHAIN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE DNA POLYMERASE TYPE-C FAMILY. DNAE
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE006285; AAK04594.1; -.
CC InterPro: IPR003141; PBP_N.
CC InterPro: IPR004805; POLC_alpha.
CC InterPro: IPR004365; tRNA_antl.
CC Pfam: PF01336; tRNA_antl; 1.
CC Pfam: PF02231; PBP_N; 1.
CC SMART: SM00481; POLITAC; 1.
CC TIGRFAMs: TIGR00594; POLC; 1.
CC Transferrase; DNA-directed DNA polymerase; DNA replication;
CC Complete proteome.
CC KW Complete proteome.
CC SEQUENCE 1060 AA; 121571 MW; 44B03F0937A8D3E6 CRC64;

Query Match 8.6%; Score 86; DB 1; Length 1060;
Best Local Similarity 21.7%; Pred. No. 5.1;
Matches 45; Conservative 31; Mismatches 73; Indels 58; Gaps 8;

QY 14 DAMANAVAIEKGRLLFAGLREMYPGALPSDANFT---RMSYGSIKGYEPDQAWYN 69
DB 743 DALENGFEIER-----PSIWLKTIQDPVKKKIRLGLAHVQGISRLAKWIV 789
QY 70 YHTTGKGVLEKQDKSDFEAVQENILDFRTKNGRAENGQHLIAFLSNNDITGG 120
DB 790 ENOPYKTLADPEVKLIPNNFHKKENILPLIQIGAFDYADSNKGKLAAYNLADHANLNIYSD 849
QY 121 DITGNSGSPV-----FDKNGRLIGLA-----PGDNWEAMSGD 153
DB 850 DIFMASSGCGFAYHAEDEYSETEKYDEKFN--LLGIGVTPPLQNLARFEGNFTPLA-- 905
QY 154 IEFEPDLQRTISVDIRYVLEFMDKWCQ 180
DB 906 -QLVKNRRMTILVEINRYTRHRTGQ 931

RESULT 6
HNOA_ARATH STANDARD; PRT; 321 AA.
ID HNOA_ARATH
AC 09SEL7; 049507;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protease HnoA, chloroplast precursor (EC 3.4.21.-).
GN HNOA OR AT4G18370 OR F28J12.30.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20081488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Sliakema W., Entlan K.-D., Terryn N.,
RA Harris B., Ansoerge W., Brandt P., Grivell L., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA Reichert B., Portellelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohnsels J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA van der Schuren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
```

RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert M., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren B., Dirse W.,
 RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetler P.,
 RA Berner S., Hempel S., Feldpausch M., Lambers S., Van den Daele H.,
 RA De Keyser A., Buysse C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
 RA Pellett A., Rajadream M.A., Lyne M., Benes V., Reckmann S.,
 RA Borrova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Faltmann B., Granderath K., Dauner D., Herzl A.,
 RA Neumann S., Argitjon A., Vitale D., Ligouri R., Piravandi E.,
 RA Massenet O., Quigley F., Schmidt W., Lecharny A., Aubourg S.,
 RA Schnabl S., Hiller R., Berger C., Montfort A., Casacuberta E.,
 RA Cheifor F., Cooke R., Berger C., Montfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenberg M., Barques M., Terol J., Torres A.,
 RA Perez-Peres A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heinen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sehon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kallick J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Mink P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Maria M., Marliessen R., McCombie W.R.,
 RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
 RA thaliana".
 RT Nature 402:769-777(1999).
 RN [13]
 RP SEQUENCE OF 72-82: 96-110; 150-159; 178-211 AND 306-320.
 RA Schubert M., Peterson U., Funk C., Haas B., Schroeder W.P.,
 RA Kieselbach T.,
 RT "The chloroplast lumen from Arabidopsis thaliana".
 RT Submitted (JUN-2001) to the SWISS-PROT data bank.
 CC -1- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 52C.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A GENE
 CC PREDICTION ERROR AND THE FUSION OF THE CDS FOR THIS PROTEIN WITH
 CC THE CDS FOR A PUTATIVE NUCLEIC ACID BINDING PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AF114386; AAF24050.1; -
 DR EMBL, AL021710; CAAL6717.1; ALT_SEQ.
 DR EMBL, AL161548; CAB78839.1; ALT_SEQ.
 DR MEROPS: S01.279; -
 DR InterPro: IPR001940; Pf0tase2C.
 DR InterPro: IPR001254; Ser-protease_Try.
 DR Pfam: PF00089; trypsin_1.
 DR PRINTS: PR00834; PROTEASES2C.
 KW Hydrolyase; Serine protease; Chloroplast; Thylakoid; Transit peptide.
 FT TRANSIT 1 26 CHLOROPLAST (POTENTIAL).
 FT CHAIN 27 71 THYLAKOID.
 FT DOMAIN 72 321 PROTEASE HHOA.
 FT ACT_SITE 145 145 POLY-GLO.
 FT ACT_SITE 186 186 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 264 264 CHARGE RELAY SYSTEM (POTENTIAL).

FT CONFLICT 40 40 R -> G (IN REF. 1).
 SQ SEQUENCE 321 AA; 34691 MW; 68DB81E0BD27A7A7 CRC64;
 Query Match
 Best Local Similarity 28.5%; Score 85.5; DB 1; Length 321;
 Matches 42; Conservative 16; Mismatches 67; Indels 27; Gaps 8;
 QY 45 DANFTMRMSYSGISIKYEP-ODGAWNYHTTGK-----GVLEKODPKSDEFAV-----Q 91
 DB 166 DAKGTRFSKRGKIVGDPPNDLAVLKEETGRELNPVLTGSNDLAVGSGFAIGPYG 225
 QY 92 ENLIDFRKKNRGY--AENGQ-LHIAFLSNDITGNGSGPYFKNGRLGL--AFNG 145
 DB 226 ENLITGVVSGGLRELIPSGKISSEAIQTDDINSNGSGPLDSYGHITGVNATFTFR 285
 QY 146 NWEAMSGDIEFEPDQRTSYD--IRYVLFMI 175
 DB 286 KSGMSGYNF-----ALPDTVATVYLYI 311
 RESULT 7
 TRYE_DROER STANDARD; PRT; 256 AA.
 AC PS4627:
 DT 01-OCT-1996 (rel. 34, Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Trypsin epsilon precursor (EC 3.4.21.4).
 GN TRY-EPILTON.
 OS Drosophila erecta (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NC NCBITaxID=7220;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wang S., Hickey D.A.,
 RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 51.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, U40653; AAA83240.1; -
 DR HSP: P00763; IDPO.
 DR MEROPS: S01.112; -
 DR P1ase: Pp00015080; Dery\Try-epsilon.
 DR InterPro: IPR001314; Chymotrypsin.
 DR Pfam: PF00089; trypsin_1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYPSIN_DOM.
 DR PROSITE: PS00134; TRYPSIN_HTS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Hydrolyase; Serine protease; zymogen; Signal; Multigene family.
 FT SIGNAL 1 22 PROBABLE.
 FT PROPEP 23 30 ACTIVATION PEPTIDE.
 FT CHAIN 31 256 TRYPSIN EPILTON.
 FT ACT_SITE 71 71 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 210 210 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 56 72 BY SIMILARITY.
 FT DISULFID 180 197 BY SIMILARITY.
 FT DISULFID 206 230 BY SIMILARITY.
 FT SITE 204 204 REQUIRED FOR SPECIFICITY (BY SIMILARITY).

Seq	SEQUENCE	256 AA:	27612 MW:	6A9193173D0E37F5 CAC64:
Query Match		8.4%;	Score 84.5;	DB 1: Length 256;
Best Local Similarity		23.1%;	Pred. No. 1.2;	
Matches	42;	Conservative	28;	Mismatches 71; Indels 41; Gaps 9;
Oy	4	VIAAARAIOADMANAAVYIEKGRLEFPAG-----LRMYPGRA-----	-LPS	44
Db	66	VITAAHCLOS -VDADDKLRVSTGYWRSGGSVHSYRSPFNHCGRNARTMYNDIALYRIS		124
Oy	45	DANFTMRMSYSIGKEYEPODGA-----WYNYHTTGKYLEKODRPSDEFAVOENILDF		98
Db	125	DLSFRRSIRAAVYIAADHNPREGATAVAVSGGTETESGSPITPDHL-----LAVDLREIVDS		178
Oy	99	RTKNN-----YGRAENGOLHIAFLNNNDINGSGSPVDFKNNRLLIGLAFDGNWEAMSGPI		154
Db	179	RCRSEFEFGYGRKIRKTMFL-CATAPKPKDACQDSGGPLV-SCGRLVGVV---SWGTCGGDGV		233
Oy	155	EF	156	
Db	234	RY	235	

RESULT 8	ID	TBPH_NEIMA	STANDARD:	PRT:	698 AA.
AC	068937;				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Transferrin-binding protein 2 precursor (TBP-2).				
GN	TBPH OR TBP2 OR NMA2025.				
OC	Neisseria meningitidis (serogroup A).				
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_Taxid=65699;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Z2491 / Serogroup A / Serotype 4a;				
RA	Brieske N., Quentin-Millet M.-J., Schenker M., Schnibbe T.,				
RA	Achtman M.;				
RT	"Human antibody responses to A and C capsular polysaccharides, IgA1				
RT	protease and transferrin-binding protein complex stimulated by				
RT	infection with Neisseria meningitidis of subgroup IV-1 or Et-37				
RT	complex."				
RL	Submitted (Apr-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Z2491 / Serogroup A / Serotype 4a;				
RX	MEDLINE=2022556; PubMed=10761919.				
RA	Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,				
RA	Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,				
RA	Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,				
RA	Jagels K., Leather S., Moulé S., Mungall K., Quail M.A.,				
RA	Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,				
RA	Whitehead S., Spratt B.G., Barrall B.G.;				
RT	"Complete DNA sequence of a serogroup A strain of Neisseria				
RT	meningitidis Z2491."				
RL	Nature 404:502-506(2000).				
CC	-1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR				
CC	TRANSFERRIN UTILIZATION (BY SIMILARITY).				
CC	-1- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid				
CC	anchor (Probable).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; AF058689; AAC13725.1; -				
CC	EMBL; AL162757; CAB85244.1; -				

DR	InterPro	IPR000437	Prok_LipoProt.
DR	InterPro	IPR001677	transferrin_Lipid
DR	Plan	PF01298	LipoProtein_5; 1
DR	ProSITE	PS00013	PROKAR_LipoPROTEIN; FALSE_NEG.
KW	Outer membrane		Receptor; Signal; LipoProtein; Complete proteome.
FT	SIGNAL	1	20
FT	CHAIN	21	698
FT	LIPID	21	21
SO	SEQUENCE	698 AA	75176 MW; 83FBE14D0F617B1P CnC64;

Query Match	8.3%	Score 83	DB 1	Length 698
Best Local Similarity	21.3%	Pred. No. 5.8		
Matches	37	Conservative 24	Mismatches 71	Indels 42
				Gaps 6
QY	2	KSVAARALQDAANNAVAIEKGRLEFAGIREMYPCGALPSDANFMRMSYSIKGYE	61	
	:	:	:	:
DB	505	KNSKAMQAGSESSQADAKTEBYGSGMFLQG--EXTDEKEIFINDDNVYR-----	552	
	:	:	:	:
QY	62	PDGAMVNYHTTGK---GVLEKQDPKSDSEFAVOENIDLFTFKNYGRYAENGQLIAFL	117	
	:	:	:	:
DB	553	---GSMYGIHANGTSMGNSADKEGGRADFTV-----NFGRIKINGTL-----	593	
	:	:	:	:
QY	118	SNNDITGNGSGSPVFDKNGRLIGLAFDGNWEMSGDIEEPDLOQTITVSDIRYV	171	
	:	:	:	:
DB	594	-----TADNRQALPTIYGVGDEIGNEFSG--AKTADSGDILDQSNNTTRPKAYI	640	
	:	:	:	:

```

RESULT 9
VPS_BPMU                                STANDARD:          PRT:          504 AA.
ID_VPS_BPMU
AC_Q9TIV0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tail fiber protein (9p5).
S OR 49.
GN Bacteriophage Mu.
OS Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC Mu-like viruses.
OX NCBI_TaxID=10677;
RN [1]
RP SEQUENCE FROM N.A.
RA Morgan G. Hatfull G., Hendrix R.;
RT "Genome of bacteriophage Mu and comparison with the Haemophilus
RL Influenzae Mu-like prophage FluM.";
RT Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC -I- SIMILARITY: SOME. TO H. INFLUENZAE H11522.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on the way
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF083977; AAF01127.1.
DR InterPro; IPR005068; Phage_fiber_2.
DR Pfam; PF03406; Phage_fiber_2; 1.
KW Fiber protein.
SQ SEQUENCE 504 AA: 55360 MW: C564B6746F53993B CRC64;

```

Query Match	8.2%	Score 82;	DB 1;	Length 504;
Best Local Similarity	19.9%;	Pred. No. 4.8;		
Matches 46;	Conservative 22;	Mismatches 71;	Indels 92;	Gaps 9;
QY	8	ARAIADAMANNYALEKRGKRLFFAGLRREYYP--GALPSDAFTWRM-----SYGSLKG	59	
Db	259	ALSYLVNIAIVNGEVIISKAN---GLRIAYGNYGFFIRINDGSNTYEMLTNSGDNNGTYNG	314	
QY	60	YEPDGAWYNYHT--TGKGLERKODKSPDEFAVQ-----EMLD--	96	

Db 315 LRP---LWLNATGAVSKRGSLVSGDRLSDRFPAINSSNGMIQMRDNNATFGKNIVNTD 371
 QY 97 -----LFTKNGRY-----AENGOLHIAFLSNNDINGSGSPVDPKNGR----- 137
 Db 372 SAQALLRQNHARKRMIGSLGKNGKFCITYMINNSRTANGIDGQAYNDNNWNLGQAYVIG 431
 QY 138 -----LIGLAFDGNWEAMSGD 153
 Db 432 NYANFDSRYRVDRVRLGTQSLTGGLSHRDYKAPSGHVITGFTHTGDMWEMQCGD 482

RESULT 10
 BCL_MOUSE STANDARD: PRT: 874 AA.
 ID BCL_MOUSE 061140; 060869;
 AC 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE CRK-associated substrate (p130Cas) (Breast cancer anti-estrogen resistance 1 protein).
 GN BCAR1 OR CRKAS OR CAS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS CAS-A AND CAS-B), AND INTERACTION WITH FOCAL ADHESION KINASE.
 RC TISSUE=Embryo;
 RX MEDLINE=9608679; PubMed=7479864;
 RA Polte T.R., Hanks S.K.;
 RT "Interaction between focal adhesion kinase and Crk-associated tyrosine kinase substrate p130Cas";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:10678-10682(1995).
 RN [2]
 RP INTERACTION WITH NEPHROCYSTIN.
 RX MEDLINE=20249316; PubMed=10739664;
 RA Donaldson J.C., Dempsey P.J., Reddy S., Boulton A.H., Coffey R.J., Hanks S.K.;
 RT "Crk-associated substrate p130Cas interacts with nephrucystin and both proteins localize to cell-cell contacts of polarized epithelial cells";
 RL Exp. Cell Res. 256:168-178(2000).
 CC -I- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION. IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY). HAS BEEN SHOWN TO BE ESSENTIAL IN CARDIOVASCULAR DEVELOPMENT DURING EMBRYOGENESIS.
 CC -I- SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1, ADAPTER PROTEIN CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL (BY SIMILARITY). INTERACTS WITH NEPHROCYSTIN AND PTK2B.
 CC -I- SUBCELLULAR LOCATION: FOCAL ADHESIONS. UNPHOSPHORYLATED FORM LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO THE MEMBRANE UPON TYROSINE PHOSPHORYLATION (BY SIMILARITY).
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS: CAS-B (SHOWN HERE) AND CAS-A; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -I- DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTRATE DOMAIN) CONTAINING MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-BINDING SITES POTENTIALLY BIND CRK, NCK AND ABL SH2 DOMAINS. THE HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOPODAL GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.
 CC -I- DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM RESPONSE ELEMENT (SRE) (BY SIMILARITY).
 CC -I- DOMAIN: THE SH3 DOMAIN IS NECESSARY FOR THE LOCALIZATION OF THE PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-RICH REGION OF FOCAL ADHESION KINASE 1.
 CC -I- PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE DYDVLH MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX (BY SIMILARITY).

CC -I- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -I- SIMILARITY: BELONGS TO THE CAS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-stb.ch/announce/> or send an email to license@isb-stb.ch).
 CC -----
 CC EMBL: U48853; AAA93381.1; -
 CC EMBL: U28151; AAA93248.1; -
 CC HSSP: P07751; 1BK2.
 CC MGD: MGI:108091; Crkas.
 CC InterPro: IPR001452; SH3.
 CC Pfam: PF00018; SH3; 1.
 CC PRINTS: PR00452; SH3DOMAIN.
 CC ProDom: PD000066; SH3; 1.
 CC SMART: SM00326; SH3; 1.
 CC PROSITE: PS50002; SH3; 1.
 CC KW Phosphorylation; SH3 domain; SH3-binding; Cell adhesion;
 CC KW Alternative splicing;
 CC FT DOMAIN 6 26 SH3.
 CC FT DOMAIN 74 87 PRO-RICH.
 CC FT DOMAIN 119 420 SUBSTRATE FOR KINASES (BY SIMILARITY).
 CC FT DOMAIN 426 618 SRR-RICH.
 CC FT SITE 639 647 SH3-BINDING (POTENTIAL).
 CC FT DOMAIN 750 800 DIVERGENT HELIX-LOOP-HELIX MOTIF.
 CC FT VARSPLIC 1 4 MTPV -> MKYL (IN ISOFORM CAS-A).
 CC SQ SEQUENCE 874 AA; 94256 MW; 5B9EDD76532BDDBA CRC64;

Query Match 8 2%; Score 82; DB 1; Length 874;
 Best Local Similarity 24.4%; Pred. No. 9.4;
 Matches 43; Conservative 17; Mismatches 60; Indels 56; Gaps 9;

QY 4 VINAARAIOADMANAYALEKGRLEFPAGLREMP-----RALP 43
 Db 583 LVACSAVPEDEKQALSLFLHGNSLFLFRRTKAPGPPESSSLHPPTDKASSIQSRPLP 642
 QY 44 SDANFTMRMSYSING-YEPDQGAW---YNY-HTTGKYLEKODPKSPDEPAVOENILDF 98
 Db 643 SPEKFT---SODSPDQYNSSEGWMEDYVHLQCKEEFEK-----TQKELLERG 690
 QY 99 RRTKNGRYAENGOLHIAFLS-----NNDITGNGSSPVFDKNGRLIGL 141
 Db 691 NTRGCK-----GQLELQQLKQFERLEQEVSRPIDHLAMWTAPQPLVP--GRTGGL 740

RESULT 11
 BCL_MOUSE STANDARD: PRT: 968 AA.
 ID BCL_MOUSE 063767; 063766;
 AC 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE CRK-associated substrate (p130Cas) (Breast cancer anti-estrogen resistance 1 protein).
 GN BCAR1 OR CRKAS OR CAS.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC TISSUE=Fibroblast;
 RX MEDLINE=94349922; PubMed=8070403;
 RA Sakai R., Iwamatsu A., Hirano N., Ogawa S., Tanaka T., Mano H., Yazaki Y., Hirai H.;
 RT "A novel signaling molecule, p130, forms stable complexes in vivo with v-Crk and v-Src in a tyrosine phosphorylation-dependent manner";
 RL EMBO J. 13:3748-3756(1994).
 RN [2]

RP TYROSINE PHOSPHORYLATION BY FOCAL ADHESION KINASE.
 RX MEDLINE=98030588; PubMed=9360983;
 RA Tachibana K., Umano T., Fujita H., Ohashi Y., Kamiguchi K., Iwata S.,
 RA Hirai H., Morimoto C.;
 RT "Tyrosine phosphorylation of Crk-associated substrates by focal
 RT adhesion kinase. A putative mechanism for the integrin-mediated
 RT tyrosine phosphorylation of Crk-associated substrates.";
 RL J. Biol. Chem. 272:29083-29090(1997).
 CC -1- FUNCTION: DOCKING PROTEIN WHICH PLAYS A CENTRAL COORDINATING ROLE
 CC FOR TYROSINE-KINASE-BASED SIGNALING RELATED TO CELL ADHESION.
 CC IMPLICATED IN INDUCTION OF CELL MIGRATION (BY SIMILARITY).
 CC -1- SUBUNIT: FORMS COMPLEXES IN VIVO WITH FOCAL ADHESION KINASE 1,
 CC ADAPTER PROTEIN CRKL AND LYN KINASE. CAN HETERODIMERIZE WITH CASL.
 CC INTERACTS WITH NEPHROCYSTIN AND PK2B (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: FOCAL ADHESIONS AND STRESS FIBERS.
 CC UNPHOSPHORYLATED FORM LOCALIZES IN THE CYTOPLASM AND CAN MOVE TO
 CC THE MEMBRANE UPON TYROSINE PHOSPHORYLATION.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM (SHOWN HERE)
 CC AND A SHORT ISOFORM, MAY BE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. HIGHER EXPRESSION IN LUNG,
 CC INTESTINE AND TESTIS.
 CC -1- DOMAIN: CONTAINS A CENTRAL DOMAIN (SUBSTRATE DOMAIN) CONTAINING
 CC MULTIPLE POTENTIAL SH2-BINDING SITES AND A C-TERMINAL DOMAIN
 CC CONTAINING A DIVERGENT HELIX-LOOP-HELIX (HLH) MOTIF. THE SH2-
 CC BINDING SITES POTENTIALLY BIND CRK, NCK AND ABL SH2 DOMAINS. THE
 CC HLH MOTIF IS ABSOLUTELY REQUIRED FOR THE INDUCTION OF PSEUDOHYPHAL
 CC GROWTH IN YEAST AND MEDIATES HETERODIMERIZATION WITH CASL.
 CC -1- DOMAIN: A SERINE-RICH REGION PROMOTES ACTIVATION OF THE SERUM
 CC RESPONSE ELEMENT (SRE).
 CC -1- DOMAIN: THE SH3 DOMAIN IS NECESSARY FOR THE LOCALIZATION OF THE
 CC PROTEIN TO FOCAL ADHESIONS AND INTERACTS WITH ONE PROLINE-RICH
 CC REGION OF FOCAL ADHESION KINASE 1.
 CC -1- PTM: FOCAL ADHESION KINASE 1 PHOSPHORYLATES THE PROTEIN AT THE
 CC YDVLH MOTIF. THE SRC FAMILY KINASES ARE RECRUITED TO THE
 CC PHOSPHORYLATED SITES AND CAN PHOSPHORYLATE OTHER TYROSINE
 CC RESIDUES. TYROSINE PHOSPHORYLATION IS TRIGGERED BY INTEGRIN
 CC MEDIATED ADHESION OF CELLS TO THE EXTRACELLULAR MATRIX.
 CC -1- DISEASE: APPEARS TO HAVE A CENTRAL FUNCTION IN TRANSFORMATION OF
 CC SOME CELL TYPES.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE CAS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D29766; BAA06169.1; -;
 DR EMBL: D29766; BAA06170.1; -;
 DR HSSP: P07751; 1BK2.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 DR Phosphorylation: SH3 domain; SH3-binding; Cell adhesion;
 KW Alternative splicing.
 FT DOMAIN 97 159 SH3.
 FT DOMAIN 168 181 PRO-RICH.
 FT DOMAIN 213 514 SUBSTRATE FOR KINASES.
 FT DOMAIN 520 712 SER-RICH.
 FT SITE 733 741 SH3-BINDING (POTENTIAL).
 FT DOMAIN 844 894 DIVERGENT HELIX-LOOP-HELIX MOTIF.
 FT VARSPLIC 98 MISSING (IN SHORT ISOFORM).
 SO SEQUENCE 968 AA; 104262 MW; B6B1641BF6D6D377 CRC64;

QY 4 VIAAARAIQADAMANAVALTEKGRLEFAGLRREMPG-----RALP 43
 DB 677 LVACSRAPVEDKQQLASFLHGASLLFRRTKAPGCGPSSSSHLNPTDKASSIOSRPLP 736
 QY 44 SDANFTMRMSYSGISG-YEPDGAW---YNY-HTTGKGVLEKQDKSPDEFVAQENILDLF 98
 DB 737 SPPKFT-----SDSPGQXENSEGWMEDYDVHLLQKEEFK-----TQKELLEKG 784
 QY 99 RIKNKGRAENQQLHIAFLS-----NNDITGNGSGSVFPDKNGLLGL 141
 DB 785 NIVROCK-----GQLRLQQLKQPERLEQEVSRPDIHDLANMTPADLPV--GRTGL 834
 RESULT 12
 MCRB_METBA
 ID MCRB_METBA STANDARD: PRT: 433 AA.
 AC P07955;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Methyl-coenzyme M reductase beta subunit (EC 1.8.-.-).
 GN MCRB.
 OS Methanosarcina barkeri.
 OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 OX NCBI_TaxID=2208;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Fusaro / DSM 804;
 RX MEDLINE=87231011; PubMed=3502709;
 RA Bokranz M., Klein A.;
 RT "Nucleotide sequence of the methyl coenzyme M reductase gene cluster
 RT from Methanosarcina barkeri."
 RL Nucleic Acids Res. 15:4350-4351(1987).
 CC -1- FUNCTION: Reduction of methyl-coenzyme M (2-(methylthio)
 CC ethanesulfonic acid) with 7-mercaptoheptanoylthreosine phosphate
 CC to methane and an heterodisulfide.
 CC -1- CATALYTIC ACTIVITY: CH(3)-S-COM + H-S-HRP = CH(4) + COM-S-S-HRP.
 CC -1- COFACTOR: THE ENZYME COMPLEX BINDS TIGHTLY (BUT NOT COVALENTLY)
 CC TO TWO MOLECULES OF COENZYME F430. F430 IS A YELLOW NICKEL
 CC PORPHINOID.
 CC -1- PATHWAY: Methanogenesis; Last step.
 CC -1- SUBUNIT: HEXAMER OF TWO ALPHA, TWO BETA, AND TWO GAMMA CHAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y00158; CAA68353.1; -;
 DR PIR: A29525; A29525.
 DR HSSP: P11560; 1MRO.
 DR InterPro: IPR003179; MCR_beta.
 DR Pfam: PF02241; MCR_beta.1.
 DR Pfam: PF02783; MCR_beta_N.1.
 DR Methanogenesis; Oxidoreductase.
 KW INIT_MET
 FT INIT_MET 0 0
 SO SEQUENCE 433 AA; 45289 MW; BABFAA3709361A9 CRC64;

Query Match 8.0%; Score 80.5; DB 1; Length 433;
 Best Local Similarity 20.1%; Pred. No. 5.4;
 Matches 56; Conservative 34; Mismatches 74; Indels 115; Gaps 13;
 QY 2 KSVIAAARAIQADAMANAVALTEKGRLEFAGLRREMPG--ALPSDANFTMRMSG----- 55
 DB 40 RSVAVNLAGIQC-ALASGKMGKGRQILGRGLINDIVGNADIAENVKLVQVDEGDDTN 98
 QY 56 --SINGYEPDGAWNYHTTGKGVLEKQDKPS-----DEFVAQENILDLFR 99
 DB 56 --SINGYEPDGAWNYHTTGKGVLEKQDKPS-----DEFVAQENILDLFR 99

```

Db 99 VIRKVG-----GKSL-IOSEKRIIAGAFMSATVGAATOTIMDFG 143
OY 100 TKNY-----GRYAE-----NGOL-----HTAPLSNND 121
Db 144 TDYDAPRYKSAVNGSYPTQIMDMGGVCGILSPONNEGSLRNIMANHVAALSNRN 203
OY 122 ITGNSGSPVFDKNG-----RLIGLAFDG-NWEAMSGDIEFEPDLQRTISV 166
Db 204 AMASALSIYEGSGIFEMGAVGMEFRHQLLGLAYOGANLNLVDIVENCKDGICT 263
OY 167 DIRVY-----LFMID-----KMGQC 181
Db 264 VIESVVRRALEAGIISVDKTAPSGNYFNKANDVPKWNAC 302

RESULT 13
STSP-STAU
ID STSP-STAU STANDARD: PRT; 336 AA.
AC P04188;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glutamy endopeptidase precursor (EC 3.4.21.19) (Staphylococcal serine
  proteinase) (V8 proteinase) (Endoproteinase Glu-C).
GN SSPA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
  SEQUENCE FROM N.A.
  POLG-KUNJM
  ID POLG-KUNJM STANDARD: PRT; 3433 AA.
  AC P14335; 082983;
  DT 01-JAN-1990 (Rel. 13, Created)
  DT 01-JAN-1990 (Rel. 13, Last sequence update)
  DT 15-JUN-2002 (Rel. 41, Last annotation update)
  DE Genome polypeptide [contains: Capsid protein C (Core protein); Matrix
    protein (Envelope protein M); Major envelope protein E; Nonstructural
    proteins NS1, NS2, NS4A and NS4B; Protease/helicase (EC 3.4.21.98)
    DE (NS3): RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)).
    Kunjin virus (strain MR61C).
    OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
    OC Flavivirus.
    OX NCBI_TaxID=11078;
    RN [1]
      SEQUENCE FROM N.A.
      MEDLINE=88089524; PubMed=2826659;
      RA Coia G., Parker M.D., Speight G., Byrne M.E., Westaway E.G.;
      RT "Nucleotide and complete amino acid sequences of Kunjin virus:
        RT definitive gene order and characteristics of the virus-specified
        RT proteins.";
      RL J. Gen. Virol. 69:1-21(1988).
      CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
        CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
        CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
        CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
          CC precursor polypeptide, commonly with Asp or Glu in the p6
          CC position, Cys or Thr in p1 and Ser or Ala in p1'.
          CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
            CC (RNA)(N).
            CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
              CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
                CC PROTEIN M AND GYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
                  CC PROTEIN C AND MRNA.
                  CC -----
                  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
                    CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
                    CC the European Bioinformatics Institute. There are no restrictions on its
                    CC use by non-profit institutions as long as its content is in no way
                    CC modified and this statement is not removed. Usage by and for commercial
                    CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
                    CC or send an email to license@isb-sib.ch).
                    CC -----
                    CC EMBL; Y00356; CAA68434.1; -
                    DR EMBL; AF309515; AAG45843.1; -
                    DR PIR; A26812; PRSASK.
                    DR MEROPS; S01.269; -
                    DR InterPro; IPR001254; Ser_proteas_V8.
                    DR InterPro; IPR001254; Ser_protease_Try.
                    DR Pfam; PF00089; trypsin; 1.

```

```

DR PRINTS: PR00839; V8PROTEASE.
DR PROSITE: PS00672; V8_HIS_1.
DR PROSITE: PS00673; V8_SER_1.
KW Hydrolyase; Serine protease; Zymogen; Signal.
FT SIGNAL 1
FT PROPEP 30
FT CHAIN 69
FT ACT_SITE 119
FT ACT_SITE 161
FT ACT_SITE 237
FT CONFLICT 109
FT CONFLICT 125
FT CONFLICT 145
FT CONFLICT 193
FT CONFLICT 229
FT CONFLICT 259
FT CONFLICT 261
SQ SEQUENCE 336 AA; 36326 MW; 8B138D0C7996AA3E CRC64;
  EFN -> OFD (IN REF. 3).
  ENV -> NEVN (IN REF. 3).

```

```

Query Match      8.0%; Score 80; DB 1; Length 336;
Best Local Similarity 31.5%; Pred. No. 4.4; Mismatches 24; Indels 0; Gaps 0;
Matches 17; Conservative 13;

```

```

OY 123 TGNSSGSPVFDKNGRLIGLAFDGNWEAMSGDIEFEPDLQRTISVDIRYVLFMID 176
Db 233 TGNSSGSPVFNENKENVYIGIMGSPVFNENKAVFTINENVRFLKQNIEDHIFAND 286
  |||||

```

```

RESULT 14
POLG-KUNJM
ID POLG-KUNJM STANDARD: PRT; 3433 AA.
AC P14335; 082983;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Genome polypeptide [contains: Capsid protein C (Core protein); Matrix
  protein (Envelope protein M); Major envelope protein E; Nonstructural
  proteins NS1, NS2, NS4A and NS4B; Protease/helicase (EC 3.4.21.98)
  DE (NS3): RNA-directed RNA polymerase (EC 2.7.7.48) (NS5)).
  Kunjin virus (strain MR61C).
  OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
  OC Flavivirus.
  OX NCBI_TaxID=11078;
  RN [1]
    SEQUENCE FROM N.A.
    MEDLINE=88089524; PubMed=2826659;
    RA Coia G., Parker M.D., Speight G., Byrne M.E., Westaway E.G.;
    RT "Nucleotide and complete amino acid sequences of Kunjin virus:
      RT definitive gene order and characteristics of the virus-specified
      RT proteins.";
      RL J. Gen. Virol. 69:1-21(1988).
      CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
        CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
        CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
        CC -1- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral
          CC precursor polypeptide, commonly with Asp or Glu in the p6
          CC position, Cys or Thr in p1 and Ser or Ala in p1'.
          CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
            CC (RNA)(N).
            CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
              CC LIPIDPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
                CC PROTEIN M AND GYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
                  CC PROTEIN C AND MRNA.
                  CC -----
                  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
                    CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
                    CC the European Bioinformatics Institute. There are no restrictions on its
                    CC use by non-profit institutions as long as its content is in no way
                    CC modified and this statement is not removed. Usage by and for commercial
                    CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
                    CC or send an email to license@isb-sib.ch).
                    CC -----
                    CC EMBL; D00246; BAA00176.1; -

```

```
DR PIR; A28697; GNMVKV.
DR HSSP; P14336; I5VB.
DR MEROPS; S07.001; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000069; Flavi_M.
DR InterPro; IPR001157; Flavi_NSL.
DR InterPro; IPR000752; Flavi_NS2A.
DR InterPro; IPR000487; Flavi_NS2B.
DR InterPro; IPR000404; Flavi_NS2A.
DR InterPro; IPR001528; Flavi_NS4B.
DR InterPro; IPR000208; Flavi_NS5.
DR InterPro; IPR001122; Flavi_capsidC.
DR InterPro; IPR000336; Flavi_glycoprote.
DR InterPro; IPR001850; Flavi_helicase.
DR InterPro; IPR002535; Flavi_propep.
DR InterPro; IPR002877; FtsJ.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00869; Flavi_glycoprote; 1.
DR Pfam; PF00948; Flavi_NSL; 1.
DR Pfam; PF00949; Flavi_helicase; 1.
DR Pfam; PF00972; Flavi_NS5; 1.
DR Pfam; PF01002; Flavi_NS2B; 1.
DR Pfam; PF01003; Flavi_capsid; 1.
DR Pfam; PF01004; Flavi_M; 1.
DR Pfam; PF01005; Flavi_NS2A; 1.
DR Pfam; PF01349; Flavi_NS4B; 1.
DR Pfam; PF01350; Flavi_NS4A; 1.
DR Pfam; PF01570; Flavi_propep; 1.
DR Pfam; PF01728; FtsJ; 1.
DR Pfam; PF02832; Flavi_glycop_C; 1.
DR ProDom; PD001496; Flavi_NSL; 1.
DR ProDom; PD001556; Flavi_glycoprote; 1.
DR SMART; SM00490; HELICC; 1.
KM Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;
KM Core protein; Coat protein; Envelope protein; Hydroxylase; Helicase;
KM ATP-binding; Transmembrane; Nonstructural protein.
FT INIT_MBT 1 1
FT CHAIN 1 125
FT PROPEP 124 215
FT CHAIN 216 290
FT CHAIN 291 791
FT CHAIN 792 1143
FT CHAIN 1144 1374
FT CHAIN 1375 1505
FT CHAIN 1506 2124
FT CHAIN 2125 2273
FT CHAIN 2274 2528
FT CHAIN 2529 3433
FT DOMAIN 388 401
FT NP_BIND 1699 1706
FT SITE 1790 1793
FT DISULFID 223 320
FT DISULFID 350 406
FT DISULFID 364 395
FT DISULFID 382 411
FT DISULFID 480 578
FT DISULFID 595 626
FT CARBOHYD 138 138
FT CARBOHYD 921 921
FT CARBOHYD 966 966
FT CARBOHYD 998 998
SQ SEQUENCE 3433 AA; 381363 MW; EE4B888A7D040B99 CRC64;
```

```
Query Match 8.0%; Score 80; DB 1; Length 3433;
Best Local Similarity 21.4%; Pred. No. 76;
Matches 43; Conservative 28; Mismatches 50; Indels 80; Gaps 10;
```

```
OY 52 MSGSITKGYEPDQGANYN-----YHTTGKYLEKODPKSDEF--AVQENIILDFRTKNY 103
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1531 MTRGLGISTQAGAGVWGVCFHTLMHTTKGAALMSGEGRLDPYWSGVKEDRL-----CY 1584
```

```
OY 104 G-----RYAENGQHLIAFL-----SNN-----DITGNSGSPV 131
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1585 GGPWKLOHKHKNQODEVQMIIVPEPKNVKNVQTPKPGVFKTPPEGELGAVYLDPEPTGSGSPI 1644
OY 132 PDKNGRLIGLAFDG---NWEAMSGDIE-----FEPDLORTISVDI----- 168
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1645 VDKNGDVIGLYNGVIMPGSTISAIVQGERMDEVVPAGFEFEMLRKQITVYLDHPGAG 1704
OY 169 --RYVLFMIDKMGCCPRLIOE 187
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1705 KTRRL-----PQIKE 1716
```

```
RESULT 15
Y030_UREPA STANDARD; PRT; 747 AA.
AC O9PRB5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein U0030.
GN U0030.
OC Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OS Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum."
RL Nature 407:757-762(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AE002102; AAF30435.1;
DR Hypothetical protein; Transmembrane; Complete proteome.
KW TRANSMEM POTENTIAL.
SQ SEQUENCE 747 AA; 85860 MW; 68FFD940C28D02AF CRC64;
```

```
Query Match 7.9%; Score 79.5; DB 1; Length 747;
Best Local Similarity 23.4%; Pred. No. 13;
Matches 39; Conservative 21; Mismatches 40; Indels 67; Gaps 8;
```

```
OY 27 RLFFPGLRMVGRRLPSDA---NFTMRMSYGS-IKGYEPDQGANYNHTTGKYLEKOD 82
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 585 QLFSGG---YGDVNPNSAIVSMRGSKYSGLQAF----- 618
OY 83 PKSDEFQVQENIILDFRTE---NYGRAENGQHLIAFLSN-----NDITGNSG 128
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 619 ---DREIKNESLIDYGGPQINNIDGYQKVGEGYLNKLFNVGTRVITSDEIDLSSGSG 675
OY 129 SPVDPKNGRLIGLAFD-----GNM-----EAMSGDIE 155
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 676 SMIDSNENLVGIIHFASLNSRAYGAPNDSMIGNLFLVAQSODLSGID 722
```

```
Search completed: December 20, 2002, 17:28:39
Job time : 37 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2002, 15:04:15 ; Search time 87 Seconds
(without alignments)
452.356 Million cell updates/sec

Title: US-10-008-355-2_COPY_522_712

Perfect score: 1003
Sequence: 1 SKSVIAARAIQADAMANAY.....LFMIDKMGCPRLIQELKLI 191

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.humano:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.verticillate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334	33.3	716	16	Q9PC94 xylella fas
2	101.5	10.1	305	16	Q8XUP6 ralstonia s
3	93	9.3	513	16	Q97LAF clostridium
4	90.5	9.0	408	12	Q96604 avian infec
5	90	9.0	711	2	Q51284 neisseria m
6	89.5	8.9	856	4	Q76022 homo sapien
7	89.5	8.9	856	4	Q9BUJ2 homo sapien
8	89.5	8.9	859	11	Q8VDM6 mus musculu
9	89.5	8.9	3411	12	Q98803 yellow feve
10	89.5	8.9	3411	12	Q91857 yellow feve
11	89.5	8.9	3411	12	Q9YMN2 yellow feve
12	89.5	8.9	3411	12	Q9YMN1 yellow feve
13	89.5	8.9	3411	12	Q9YMN0 yellow feve
14	89.5	8.9	3411	12	Q89275 yellow feve
15	89.5	8.9	3411	12	Q89277 yellow feve
16	89.5	8.9	3411	12	Q89278 yellow feve

17	89.5	8.9	3411	12	Q89276	Q89276 yellow feve
18	86.5	8.6	576	2	Q9RF25	Q9RF25 campylobact
19	86.5	8.6	3411	12	Q9YRV3	Q9YRV3 yellow feve
20	86	8.6	281	17	Q8TL57	Q8TL57 methanosarc
21	85.5	8.5	323	10	Q932W5	Q932W5 arabidopsis
22	84.5	8.4	576	2	Q85183	Q85183 campylobact
23	84	8.4	238	2	Q9FDD7	Q9FDD7 staphylococ
24	84	8.4	729	16	Q8YC77	Q8YC77 brucella me
25	83.5	8.3	1004	15	Q994B3	Q994B3 human immun
26	83	8.3	263	16	Q8RGN9	Q8RGN9 fusobacteri
27	83	8.3	316	2	Q9FBG1	Q9FBG1 staphylococ
28	83	8.3	398	11	Q8VE75	Q8VE75 mus musculu
29	83	8.3	682	2	Q9EXB5	Q9EXB5 neisseria m
30	83	8.3	684	2	Q9JPM0	Q9JPM0 neisseria m
31	83	8.3	699	2	Q9EXC4	Q9EXC4 neisseria m
32	82.5	8.2	726	2	Q30713	Q30713 flavobacter
33	82.5	8.2	759	16	Q92A91	Q92A91 listeria in
34	82	8.2	402	16	Q9AB91	Q9AB91 caulobacter
35	82	8.2	798	10	Q9LIS0	Q9LIS0 nicotiana r
36	82	8.2	839	5	Q8R610	Q8R610 dictyosteli
37	81.5	8.1	409	12	Q96605	Q96605 avian infec
38	81.5	8.1	493	16	Q98PH9	Q98PH9 mycoplasma
39	81.5	8.1	549	10	Q93Z39	Q93Z39 arabidopsis
40	81.5	8.1	603	10	Q9SZT9	Q9SZT9 arabidopsis
41	81	8.1	239	2	Q9KH49	Q9KH49 staphylococ
42	81	8.1	239	16	Q53782	Q53782 staphylococ
43	81	8.1	283	2	Q9F474	Q9F474 ehlichia c
44	81	8.1	663	16	Q8ZNL0	Q8ZNL0 salmonella
45	81	8.1	663	16	Q8Z597	Q8Z597 salmonella

ALIGNMENTS

RESULT 1	ID	Q9PC94	PRELIMINARY:	PRT:	716 AA.
AC	Q9PC94:				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DE	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Hypothetical protein Xf1887.				
GN	Xf1887.				
OS	Xylella fastidiosa.				
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;				
OC	Xylella.				
NC	NCBI_TaxID=2371;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=945C;				
RX	MEDLINE=20365717; PubMed=10910347;				
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,				
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,				
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Bionesi M.R.S.,				
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,				
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa Neto C.M.,				
RA	Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,				
RA	Faciuncani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,				
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furian L.R.,				
RA	Gartier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,				
RA	Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,				
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,				
RA	Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,				
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Martins C.L.,				
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuna A.Y.,				
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,				
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,				
RA	Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,				
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,				
RA	Pelietro B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,				
RA	Queiro R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,				
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,				
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,				

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen *Xylella fastidiosa*,"
 RL Nature 406:151-159(2000).
 DR EMBL: AE004008; AAF84693.1; -.
 DR MEROPS: S46.001; -.
 KW Hypothetical protein: Complete proteome.
 SQ SEQUENCE 716 AA; 79375 MW; EAF086E2315BBD5C CRC64;

Query Match 33.3%; Score 334; DB 16; Length 716;
 Best Local Similarity 40.0%; Pred. No. 5,1e-22;
 Matches 74; Conservative 40; Mismatches 53; Indels 18; Gaps 5;

OY 8 ARAIADANANVAIEKGRLEFAGLREMYPGRALPDPANFTMRMSYGSIGYEDPGAM 67
 DB 546 ARPIYLOALAD-YNKSHGK-----FYYP-----DANSSLRTTFGVKYSRKGVE 590
 OY 68 YVHTTGKGVLEKODPKSDEFAVQENIDLFRTKNYGRYAEN--QQLATLPLSNDITGG 125
 DB 591 YPFTLLGCVMAK-NMGVFPDPSKRLINAKASYANLADRGIVPVPNFLSDITGG 649
 OY 126 NGSPPYFKNGSLIGLAFPGNWEAMSGDIEFPDLORTISVIRYVLEMDKMGCCPRLI 185
 DB 650 NSGSPVLAHGRLVGLAFDGNWESVSNMVPDVRTIADVSRYVOWIMTEVAPPHLL 709
 OY 186 QELKL 190
 DB 710 KELNL 714

RESULT 2

O8XUP6 PRELIMINARY; PRT; 305 AA.
 AC 08XUP6.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Putative serine protease protein (EC 3.4.21.-).
 GN RSC3140 OR RS00463.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-GM1000;
 RX MEDLINE-21681879, PubMed-11823852;
 RA Salanoubat M., Genin S., Attiguenave F., Gouzy J., Mangenot S.,
 RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sautier P., Thebault P., Whalen M., Winkler P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen *Ralstonia solanacearum*,"
 RL Nature 415:497-502(2002).
 DR EMBL: AL646073; CAD16849.1; -.
 DR InterPro: IPR000104; Antifreeze_1.
 DR InterPro: IPR001254; Ser_protease_Try.
 DR Pfam: PF00089; trypsin_1.
 DR PRINTS: PR00308; ANTI-FREEZE1.
 KW Hydrolyse: Complete proteome.
 SQ SEQUENCE 305 AA; 32269 MW; 284EC5874BF94327 CRC64;

Query Match 10.1%; Score 101.5; DB 16; Length 305;
 Best Local Similarity 24.1%; Pred. No. 0.32;
 Matches 33; Conservative 24; Mismatches 51; Indels 29; Gaps 3;

OY 16 MANAAIEKGRLEFAGLREMYPGRALPDPANFTMRMSYGSIGK-----GYEPD 64
 DB 43 ITNCVLAHGKGVMAK-----RGNANFGARLQIPYVERDLQLRVADFTHPVPT 91

OY 65 GAWNYHTTGKGVLEKODPKSDEFAVQENIDLFRTKNYGRYAENGOLHIAFLSNDITG 124
 DB 92 LAPGALVTVGQVVAIGNPLGELITSEGLISLRTDDGRLLK-----SVQTSALSR 144
 OY 125 GNSGSPVDFKNRRLGL 141
 DB 145 GSSGGGLFDANRRLGI 161

RESULT 3

O97LA8 PRELIMINARY; PRT; 513 AA.
 AC O97LA8;
 DT 01-OCT-2001 (TREMBLrel. 18, Created)
 DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Sensory transduction histidine kinase.
 GN CAC0634.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Clostridiales; Clostridiaceae; Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE-21359325, PubMed-11466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Oiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 bacterium *Clostridium acetobutylicum*,"
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL: AE007581; AA878631.1; -.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR004358; Bact_sens_pr_C.
 DR InterPro: IPR003660; HAM.
 DR InterPro: IPR004361; His_kinA.
 DR InterPro: IPR004359; His_kinA.
 DR Pfam: PF00672; HAM_p. 1.
 DR Pfam: PF02518; HAM_p. 1.
 DR Pfam: PF00512; signal_1.
 DR PRINTS: PR00344; BCTRLSENSOR.
 DR SMART: SM00387; HAM_p. 1.
 DR SMART: SM00388; HisKA. 1.
 KW Kinase; Complete proteome.
 SQ SEQUENCE 513 AA; 59185 MW; A2F61615182423B3 CRC64;

Query Match 9.3%; Score 93; DB 16; Length 513;
 Best Local Similarity 24.7%; Pred. No. 3.9;
 Matches 37; Conservative 29; Mismatches 56; Indels 28; Gaps 7;

OY 48 FTMRMSYGSIGYEDPGAMVYHT-----TGKGVLEKODPKSDEFAVQENIDLF 98
 DB 96 FVEKKGGVIVYKTEKRYQNTLETIFNNGKVITSLP--DDYS-EDWLSCY 152
 OY 99 RTKNYGRYAENGOLHIAFLSNDITGNSGSPVDFKNRRLGL--AFDGNWEAMSGDI-- 154
 DB 153 GTRNL-KYKSGQITIKLLKNNMSG--KLSTVIDKNSVIRIQKIDSVYDKMGELSN 209
 OY 155 -----FEFPDLORTISVIRYVLEMI 175
 DB 210 NKGFKMTSGEFAAVMVLRYRFYIIFLI 239

RESULT 4

O96604 PRELIMINARY; PRT; 408 AA.
 AC O96604;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Nucleocapsid protein.

OS Avian infectious bronchitis virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus.
OX NCBI_TaxId=11120;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O3/88;
RX MEDLINE=97124667; PubMed=8955062;
RA Sapets S.I., Ashton F., Wright P.J., Ignjatovic J.;
RT "Novel variation in the N protein of avian infectious bronchitis
virus.";
RL Virology 226:412-417(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O3/88;
RA Sapets S.I., Ashton F., Wright P.J., Ignjatovic J.;
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; U52600; AAB48161.1;-
DR InterPro; IPR001218; Corona_nucleocap.
DR Pfam; PF00937; Corona_nucleoca; 1.
SQ SEQUENCE 408 AA; 45426 MW; BCFCA3DAE75D13 CRC64;

Query Match 9.0%; Score 90.5; DB 12; Length 408;
Best Local Similarity 25.9%; Pred. No. 4.9;
Matches 37; Conservative 17; Mismatches 44; Indels 45; Gaps 7;

OY 25 GKLPFAGLRE-----MYPRALPSDANFTMRMSYGSIK-----GYEPQDGA 66
DB 32 GQASNFQSLKEKKRTGEPFEGSGVPDINSYKPPQHGTYMKRORYKSGKGRKPVADA 91
OY 67 WYNYHT-TGK-GVLEKQDKPSDEFAVOENILDLFRKNYG-----RYAENGOL 112
DB 92 WYFYTGCTGPFGLDLEMGDNDVYWKAKGADTSKIGNYGVDPDKFDQAPLFTIEGG-- 149
OY 113 HTAFLSNN-----DITGNSGS 129
DB 150 -----PNNRMDFTALSIRGRNS 167

RESULT 5

OS1284
ID 051284 PRELIMINARY; PRT; 711 AA.
AC 051284;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Transferrin-binding protein 2 precursor.
GN TBP2.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B2163;
RA Legrain M., Findeli A., Villevet D., Quentin-Millet M., Jacobs E.;
RT "Molecular characterization of hybrid transferrin-binding protein 2's
from Neisseria meningitidis";
RL Submitted (AUG-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z50731; CA930398.1;-
DR InterPro; IPR001677; transferrin_bind.
DR Pfam; PF01298; Lipoprotein_5; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 20 POTENTIAL.
FT SEQUENCE 711 AA; 77057 MW; 9BC8A1671F6991D0 CRC64;

Query Match 9.0%; Score 90; DB 2; Length 711;
Best Local Similarity 21.8%; Pred. No. 12;
Matches 38; Conservative 28; Mismatches 66; Indels 42; Gaps 7;

OY 2 KSVIAAARAIQADANAIAVAIEKGRFLFAGLRPMYGRALPSDANFTMRMSYGSIKGYE 61
DB 518 KNSKSMQAGESSQADANTEQVQSGSMFLQG--ERTDEKEIPSEQNIIVYR----- 565

OY 62 PODGAWYNY---HTTGG-VLEKQDKPSDEFAVOENILDLFRKNYGRYAENGOLHIAFL 117
DB 566 ---GSNYGHIASTSMGASNAKDEKGNRAEFTY-----NFGEKKITGTL----- 606
OY 118 SNNDITGNSGSPVDFKNGRLIGLAFDGNWEAMSGDIEEPDLQRTISYDIRV 171
DB 607 -----TAENRQATFTIDKIEGNGFSGT--AKTAELGFDLDQKNTTRPKAYI 653

RESULT 6

OS76022
ID 076022 PRELIMINARY; PRT; 856 AA.
AC 076022;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE E1B-55kDa-associated protein.
GN E1B-Ap5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98406198; PubMed=9733834;
RA Gabler S., Schuett H., Groitl P., Wolf H., Shenk T., Dobner T.;
RT "E1B-55kDa-ton-associated protein: A cellular protein with RNA-
binding activity implicated in nucleocytoplasmic transport of
RT adenovirus and cellular mRNAs.";
RL J. Virol. 72:7960-7971(1998).
DR EMBL; AJ007509; CAA07548.1;-
DR InterPro; IPR003034; SAP.
DR InterPro; IPR003878; SPRY_domain.
DR InterPro; IPR003877; SPRY_receptor.
DR Pfam; PF02037; SAP; 1.
DR Pfam; PF00622; SPRY; 1.
DR SMART; SM00513; SAP; 1.
DR SMART; SM00449; SPRY; 1.
SQ SEQUENCE 856 AA; 95809 MW; 937D6ACD1BD45DFF CRC64;

Query Match 8.9%; Score 89.5; DB 4; Length 856;
Best Local Similarity 25.1%; Pred. No. 17;
Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;

OY 19 AVAIEKGRFLFAGLRPMYGRALPSDA--NFTMR-----SYGSIKGYEPQDGAWYNYH 71
DB 249 SYGVGRGRVCFEMKINEISVXKHLSTEPDPHYVRIGMSLDCSQQLGGEPPS---YGYG 305
OY 72 TTGKGVLEKQDKPSDEFAVOENILDLFRKNYG-RAENGOLHIAFLSNNDITG----- 124
DB 306 GTGK-----KSTN-----SRPENYGDKEFA-----NDVIYGCFADE 336
OY 125 -GNSGSPVDFKNGRLIGLAFDGNWEAMSGDIEEPDLQRTISYDIRV 170
DB 337 GNDVDELSTFKNGKMMGIAFRQKEALGQALYPHLYVANCVAEENF 383

RESULT 7

OS9BUJ2
ID 09BUJ2 PRELIMINARY; PRT; 856 AA.
AC 09BUJ2; 09UG75;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE E1B-55kDa-associated protein 5.
GN DKFZP586D0920.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.

RC TISSUE-PLACENTA;
RA Strusberg R.;
RL Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 355-856 FROM N.A.
RC TISSUE-UTERUS;
RA Mamult R., Heubner D., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC002564; AAH02564.1; -
DR EMBL: AL050146; CAB43291.1; -
DR InterPro: IPR002965; P_Rich_extensn.
DR InterPro: IPR003034; SAP.
DR InterPro: IPR003878; SPRY_domain.
DR InterPro: IPR003877; SPRY_receptor.
DR Pfam: PF02037; SAP; 1.
DR Pfam: PF00622; SPRY; 1.
DR PRINTS: PR01217; PRICHEXTENS.
DR SMART: SM00513; SAP; 1.
DR SMART: SM00449; SPRY; 1.
DR Hypothetical protein.
SQ SEQUENCE 856 AA; 95784 MW; AB9768B74A5F47AD CRC64;

Query Match 8.9%; Score 89.5; DB 4; Length 856;
Best Local Similarity 25.1%; Pred. No. 17;
Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;

OY 19 AYAIEGKRLFFAGLREMPGRALPSDA--NFTMR-----SYGSIKGYEPDQAMYNH 71
DB 249 SYGVRGRGRCFEMKINEISVHKLPSTEDPHVVRIGMSLDSCSTQLGEEPPS---YGYG 305
OY 72 TTGKGVLEKODPKSDEFAVOENILDLFTKNG-RYAENGQLHIAFLSNDDITG----- 124
DB 306 GCGK-----KSTN-----SRENYGDKFAE-----NDVIGCFADPE 336
OY 125 -GNSSPVFDKNGRLGLAFDGNMEAMSGDIEFEDLQRTISVDIRY 170
DB 337 CGNDVELSFTKNGKMGIAFRIQKEALGQALYPHLVKNCAVERNF 383

RESULT 8
OY 08VDM6 PRELIMINARY; PRT; 859 AA.
AC 08VDM6;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to E1B-55 kDa-associated protein 5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strusberg R.;
RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: BC021506; AAH21506.1; -
DR InterPro: IPR003034; SAP.
DR InterPro: IPR003878; SPRY_domain.
DR InterPro: IPR003877; SPRY_receptor.
DR Pfam: PF02037; SAP; 1.
DR Pfam: PF00622; SPRY; 1.
DR SMART: SM00513; SAP; 1.
DR SMART: SM00449; SPRY; 1.
SQ SEQUENCE 859 AA; 96002 MW; 197328B681DF260E CRC64;

Query Match 8.9%; Score 89.5; DB 11; Length 859;
Best Local Similarity 25.1%; Pred. No. 17;
Matches 42; Conservative 22; Mismatches 56; Indels 47; Gaps 8;

OY 19 AYAIEGKRLFFAGLREMPGRALPSDA--NFTMR-----SYGSIKGYEPDQAMYNH 71
DB 250 SYGVRGRGRCFEMKINEISVHKLPSTEDPHVVRIGMSLDSCSTQLGEEPPS---YGYG 306

OY 72 TTGKGVLEKODPKSDEFAVOENILDLFTKNG-RYAENGQLHIAFLSNDDITG----- 124
DB 307 GCGK-----KSTN-----SRENYGDKFAE-----NDVIGCFADPE 337
OY 125 -GNSSPVFDKNGRLGLAFDGNMEAMSGDIEFEDLQRTISVDIRY 170
DB 338 CGNDVELSFTKNGKMGIAFRIQKEALGQALYPHLVKNCAVERNF 384

RESULT 9
OY 098803 PRELIMINARY; PRT; 3411 AA.
AC 098803;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Polyprotein.
OS Yellow fever virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=85-82H IVORY COAST;
RA Pisano M.R., Toulou H., Nicoli J.;
RT "Homogeneity of yellow fever virus strains isolated during an epidemic
RT and post epidemic period in west Africa."
RL Submitted (APR-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U54798; AAA9612.1; -
DR HSSP: P14336; 1SVB.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001122; Flavi_capsidC.
DR InterPro: IPR000336; Flavi_glycoprote.
DR InterPro: IPR001850; Flavi_helicase.
DR InterPro: IPR000069; Flavi_M.
DR InterPro: IPR001157; Flavi_NSI.
DR InterPro: IPR000752; Flavi_NS2A.
DR InterPro: IPR000487; Flavi_NS2B.
DR InterPro: IPR000404; Flavi_NS4A.
DR InterPro: IPR001528; Flavi_NS4B.
DR InterPro: IPR000208; Flavi_NS5.
DR InterPro: IPR002535; Flavi_NS5.
DR InterPro: IPR002877; FcSj.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR001865; Ribosomal_S2.
DR Pfam: PF01003; Flavi_capsid; 1.
DR Pfam: PF00869; Flavi_glycoprot; 1.
DR Pfam: PF02832; Flavi_glycop_C; 1.
DR Pfam: PF00949; Flavi_helicase; 1.
DR Pfam: PF01004; Flavi_M; 1.
DR Pfam: PF00948; Flavi_NSI; 1.
DR Pfam: PF01005; Flavi_NS2A; 1.
DR Pfam: PF01002; Flavi_NS2B; 1.
DR Pfam: PF01350; Flavi_NS4A; 1.
DR Pfam: PF01349; Flavi_NS4B; 1.
DR Pfam: PF00972; Flavi_NS5; 1.
DR Pfam: PF01570; Flavi_propep; 1.
DR Pfam: PF01728; FcSj; 1.
DR Pfam: PF00271; helicase_C; 1.
DR ProDom: PD001496; Flavi_NSI; 1.
DR ProDom: PD001556; Flavi_glycoprote; 1.
DR SMART: SM00490; HELICC; 1.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW ATP-binding; Helicase.
SQ SEQUENCE 3411 AA; 379166 MW; A7C46E52D7E3F6F8 CRC64;

Query Match 8.9%; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%; Pred. No. 1,1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

OY 19 AYAIEGKRLFFAGLREMPGRALPSDANFTMR--SYGSIKGYEPDQAMYNH----- 71
DB 1542 AFLVNRGKRL-----IPSWASVKEDLVAWG---GSKWLEGRWDGEEEVQLIA 1585

Best Local Similarity 25.5%, Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;
Qy 19 AYAIEKGRLEFAGLRREMPGRALPSDANFTMRM-SYGSIKGYEPDQAMYNH----- 71
Db 1542 AFLVNRNGKKL-----IPSMASVKEDLVAIG---GSMKLEGRMGDEEYVQLIA 1585
Qy 72 -TTGKGVLEKODPKSDEFAVOENILDFRTKNYGRYAENGOLHIAFLSNNDITGCGSSGP 130
Db 1586 AVPGKNVNVQTKPS-----LFKVRNGG---EIGAVAL-----DYPGCTSGSP 1625
Qy 131 VFDKNGRLIGLAFDGNWEMSGDIEFPDLORT 163
Db 1626 IVNRNGEVIGLYGNG---TLVGDNSFVSASISOT 1655

RESULT 12
Q9YWN1 PRELIMINARY; PRT: 3411 AA.
AC Q9YWN1;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Polyprotein.
OS Yellow fever virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17D-204-SOUTH AFRICA VACCINE;
RX MEDLINE=98378039; PubMed=9714237;
RA Xie H., Ryman K.D., Campbell G.A., Barrett A.D.T.;
RT "Mutation in NS5 protein attenuates mouse neurovirulence of yellow fever 17D vaccine virus."
RL J. Gen. Virol. 79:1895-1899(1998).
DR EMBL: AF052445; AAC35907.1; -.
DR HSSP: P1436; 1SVB.
DR MEROPS: S07.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001122; Flav1_capsidC.
DR InterPro: IPR000336; Flav1_glycoprote.
DR InterPro: IPR001850; Flav1_helicase.
DR InterPro: IPR000069; Flav1_M.
DR InterPro: IPR001157; Flav1_NS1.
DR InterPro: IPR000752; Flav1_NS2A.
DR InterPro: IPR000487; Flav1_NS2B.
DR InterPro: IPR000404; Flav1_NS4A.
DR InterPro: IPR001528; Flav1_NS4B.
DR InterPro: IPR002038; Flav1_NS5.
DR InterPro: IPR002535; Flav1_propep.
DR InterPro: IPR002877; FtsJ.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR001092; HUH_basic.
DR InterPro: IPR001865; Ribosomal_S2.
DR Pfam: PF01003; Flav1_capsid. 1.
DR Pfam: PF00869; Flav1_glycoprote. 1.
DR Pfam: PF02832; Flav1_glycoprote. 1.
DR Pfam: PF00949; Flav1_helicase. 1.
DR Pfam: PF01004; Flav1_M. 1.
DR Pfam: PF00948; Flav1_NS1. 1.
DR Pfam: PF01005; Flav1_NS2A. 1.
DR Pfam: PF01002; Flav1_NS2B. 1.
DR Pfam: PF01350; Flav1_NS4A. 1.
DR Pfam: PF01349; Flav1_NS4B. 1.
DR Pfam: PF00972; Flav1_propep. 1.
DR Pfam: PF01570; Flav1_capsid. 1.
DR Pfam: PF01728; FtsJ. 1.
DR Pfam: PF00271; helicase_C. 1.
DR Pfam: PF00972; Flav1_NS5A. 1.
DR Pfam: PF01570; Flav1_propep. 1.
DR Pfam: PF01728; FtsJ. 1.
DR PROSITE: PS00490; HELIC_C. 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.

DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW ATP-binding; Helicase.
SQ SEQUENCE 3411 AA; 379556 MW; 1AA3ED795108EABC CRC64;
Query Match 8.9%; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%, Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;
Qy 19 AYAIEKGRLEFAGLRREMPGRALPSDANFTMRM-SYGSIKGYEPDQAMYNH----- 71
Db 1542 AFLVNRNGKKL-----IPSMASVKEDLVAIG---GSMKLEGRMGDEEYVQLIA 1585
Qy 72 -TTGKGVLEKODPKSDEFAVOENILDFRTKNYGRYAENGOLHIAFLSNNDITGCGSSGP 130
Db 1586 AVPGKNVNVQTKPS-----LFKVRNGG---EIGAVAL-----DYPGCTSGSP 1625
Qy 131 VFDKNGRLIGLAFDGNWEMSGDIEFPDLORT 163
Db 1626 IVNRNGEVIGLYGNG---TLVGDNSFVSASISOT 1655

RESULT 13
Q9YWN0 PRELIMINARY; PRT: 3411 AA.
AC Q9YWN0;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Polyprotein.
OS Yellow fever virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC Flavivirus.
OX NCBI_TaxID=11089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17D-204-SOUTH AFRICA VACCINE;
RX MEDLINE=98378039; PubMed=9714237;
RA Xie H., Ryman K.D., Campbell G.A., Barrett A.D.T.;
RT "Mutation in NS5 protein attenuates mouse neurovirulence of yellow fever 17D vaccine virus."
RL J. Gen. Virol. 79:1895-1899(1998).
DR EMBL: AF052446; AAC35908.1; -.
DR HSSP: P1436; 1SVB.
DR MEROPS: S07.001; -.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001122; Flav1_capsidC.
DR InterPro: IPR000336; Flav1_glycoprote.
DR InterPro: IPR001850; Flav1_helicase.
DR InterPro: IPR000069; Flav1_M.
DR InterPro: IPR001157; Flav1_NS1.
DR InterPro: IPR000752; Flav1_NS2A.
DR InterPro: IPR000487; Flav1_NS2B.
DR InterPro: IPR000404; Flav1_NS4A.
DR InterPro: IPR001528; Flav1_NS4B.
DR InterPro: IPR002038; Flav1_NS5.
DR InterPro: IPR002535; Flav1_propep.
DR InterPro: IPR002877; FtsJ.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR001865; Ribosomal_S2.
DR Pfam: PF01003; Flav1_capsid. 1.
DR Pfam: PF00869; Flav1_glycoprote. 1.
DR Pfam: PF02832; Flav1_glycoprote. 1.
DR Pfam: PF00949; Flav1_helicase. 1.
DR Pfam: PF01004; Flav1_M. 1.
DR Pfam: PF00948; Flav1_NS1. 1.
DR Pfam: PF01005; Flav1_NS2A. 1.
DR Pfam: PF01002; Flav1_NS2B. 1.
DR Pfam: PF01350; Flav1_NS4A. 1.
DR Pfam: PF01349; Flav1_NS4B. 1.
DR Pfam: PF00972; Flav1_propep. 1.
DR Pfam: PF01570; Flav1_capsid. 1.
DR Pfam: PF01728; FtsJ. 1.
DR PROSITE: PS00271; helicase_C. 1.

DR ProDom: PP001496; Flavi_NSI; 1.
DR ProDom: PP001556; Flavi_glycoprote; 1.
DR SMART: SM00490; HELIC; 1.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_1.
KW ATP-binding; Helicase.
SQ SEQUENCE 3411 AA; 379500 MW; 8A9B127F6623A933 CRC64;
Query Match 3.9%; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%; Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;
OY 19 AYAIEKGRLEFAGLRREMPGRALPSDANFTWRM-SYGSIKGYEPQDGAWYVH----- 71
DB 1542 AFLVNRGKKL-----IPSMASVKEDLVAYG---GSMKLEGRWDEEEVQLIA 1585
OY 72 -TTGKGVLEKDPKSDDEFAVQENILDFRTKNYGRYAENGQHLIAFLSNNDITGNSGSP 130
DB 1586 AVPGKNVNVVQTKPS-----LFKVRNGG---EIGAVAL-----DYPSTGSGSP 1625
OY 131 VFDKNGRLGLAFDGNWEMSGDIEFEPDLQRT 163
DB 1626 IVNRNGEVIGLYGNG---ILVGDNSFVSAISQT 1655
RESULT 14
O89275 PRELIMINARY; PRT; 3411 AA.
AC O89275;
DT 01-NOV-1996 (TREMBLrel, 01, Created)
DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
DE Polyprotein.
OS Yellow fever virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC NCBI_TaxID=11089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VACCINE STRAIN 17DD;
RX MEDLINE=95274286; PubMed=7754673;
RA dos Santos C.N., Post P.R., Carvalho R., Ferreira I.I., Rice C.M.,
RA Galler R.;
RT "Complete nucleotide sequence of yellow fever virus vaccine strains
RT 17DD and 17D-213.";
RL Virus Res. 35:35-41(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-VACCINE STRAIN 17DD;
RA Galler R.;
RT Submitted (NOV-1994) to the EMBL/Genbank/DBJ databases.
DR EMBL: U17066; AAC54267.1; -.
DR HSSP: P14336; ISVB.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001122; Flavi_capsidC.
DR InterPro: IPR000336; Flavi_glycoprote.
DR InterPro: IPR001850; Flavi_helicase.
DR InterPro: IPR000069; Flavi_M.
DR InterPro: IPR001157; Flavi_NSI.
DR InterPro: IPR000752; Flavi_NS2A.
DR InterPro: IPR000487; Flavi_NS2B.
DR InterPro: IPR000404; Flavi_NS4A.
DR InterPro: IPR001528; Flavi_NS4B.
DR InterPro: IPR000208; Flavi_NS5.
DR InterPro: IPR002535; Flavi_propep.
DR InterPro: IPR002877; FtsJ.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR001865; Ribosomal_S2.
DR Pfam: PF01003; Flavi_capsid; 2.
DR Pfam: PF00869; Flavi_glycoprote; 2.
DR Pfam: PF02832; Flavi_glycoprote; 2.
DR Pfam: PF00949; Flavi_helicase; 2.
DR Pfam: PF01004; Flavi_M; 2.
DR Pfam: PF00948; Flavi_NSI; 2.

DR Pfam: PF01005; Flavi_NS2A; 2.
DR Pfam: PF01002; Flavi_NS2B; 2.
DR Pfam: PF01350; Flavi_NS4A; 2.
DR Pfam: PF01349; Flavi_NS4B; 2.
DR Pfam: PF00972; Flavi_NS5; 2.
DR Pfam: PF01570; Flavi_propep; 2.
DR Pfam: PF01728; FtsJ; 2.
DR Pfam: PF00271; Helicase_C; 2.
DR ProDom: PP001496; Flavi_NSI; 1.
DR ProDom: PP001556; Flavi_glycoprote; 1.
DR SMART: SM00487; DEXC; 1.
DR SMART: SM00490; HELIC; 2.
DR PROSITE: PS00962; RIBOSOMAL_S2_1; UNKNOWN_2.
KW ATP-binding; Helicase.
FT CHAIN 2 121 CAPSID PROTEIN.
FT CHAIN 211 285 M PROTEIN.
FT CHAIN 286 778 ENVELOPE PROTEIN.
FT CHAIN 779 1187 NS1.
FT CHAIN 1188 1354 POTENTIAL.
FT CHAIN 1355 1483 POTENTIAL.
FT CHAIN 1484 2106 NS3.
FT CHAIN 2108 2394 POTENTIAL.
FT CHAIN 2395 2506 POTENTIAL.
FT CHAIN 2507 2507 NS5.
SQ SEQUENCE 3411 AA; 379429 MW; 0E7C8D189524790B CRC64;
Query Match 8.9%; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%; Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;
OY 19 AYAIEKGRLEFAGLRREMPGRALPSDANFTWRM-SYGSIKGYEPQDGAWYVH----- 71
DB 1542 AFLVNRGKKL-----IPSMASVKEDLVAYG---GSMKLEGRWDEEEVQLIA 1585
OY 72 -TTGKGVLEKDPKSDDEFAVQENILDFRTKNYGRYAENGQHLIAFLSNNDITGNSGSP 130
DB 1586 AVPGKNVNVVQTKPS-----LFKVRNGG---EIGAVAL-----DYPSTGSGSP 1625
OY 131 VFDKNGRLGLAFDGNWEMSGDIEFEPDLQRT 163
DB 1626 IVNRNGEVIGLYGNG---ILVGDNSFVSAISQT 1655
RESULT 15
O89277 PRELIMINARY; PRT; 3411 AA.
AC O89277;
DT 01-NOV-1996 (TREMBLrel, 01, Created)
DT 01-DEC-2001 (TREMBLrel, 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
DE Polyprotein.
OS Yellow fever virus.
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC NCBI_TaxID=11089;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FRENCH NEUTROTROPIC VIRUS;
RX MEDLINE=96068808; PubMed=7595382;
RA Wang E., Ryman K.D., Jennings A.D., Wood D.J., Taffs F., Minor P.D.,
RA Sanders P.G., Barrett A.D.;
RT "Comparison of the genomes of the wild-type French viscerotropic
RT strain of yellow fever virus with its vaccine derivative French
RT neurotropic vaccine.";
RL J. Gen. Virol. 76:2749-2755(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-FRENCH NEUTROTROPIC VIRUS;
RX Wang E.;
RT Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL: U21055; AAG9712.1; -.
DR HSSP: P14336; ISVB.
DR InterPro: IPR001410; DEAD.

```
DR InterPro; IPR001122; Flavi_capsidC.
DR InterPro; IPR000336; Flavi_glycoprote.
DR InterPro; IPR001850; Flavi_helicase.
DR InterPro; IPR000069; Flavi_M.
DR InterPro; IPR001157; Flavi_NS1.
DR InterPro; IPR000752; Flavi_NS2A.
DR InterPro; IPR000487; Flavi_NS2B.
DR InterPro; IPR000404; Flavi_NS4A.
DR InterPro; IPR001528; Flavi_NS4B.
DR InterPro; IPR000208; Flavi_NS5.
DR InterPro; IPR002535; Flavi_propep.
DR InterPro; IPR002877; FtsJ.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001865; Ribosomal_S2.
DR Pfam; PF01003; Flavi_capsid_2.
DR Pfam; PF00869; Flavi_glycoprot_2.
DR Pfam; PF02832; Flavi_glycop_C; 2.
DR Pfam; PF00949; Flavi_helicase; 2.
DR Pfam; PF00949; Flavi_M; 2.
DR Pfam; PF00948; Flavi_NS1; 2.
DR Pfam; PF00948; Flavi_NS1; 2.
DR Pfam; PF01005; Flavi_NS2A; 2.
DR Pfam; PF01002; Flavi_NS2B; 2.
DR Pfam; PF01350; Flavi_NS4A; 2.
DR Pfam; PF01349; Flavi_NS4B; 2.
DR Pfam; PF00972; Flavi_NS5; 2.
DR Pfam; PF01570; Flavi_propep; 2.
DR Pfam; PF01728; FtsJ; 2.
DR Pfam; PF00271; helicase_C; 2.
DR ProDom; PD001496; Flavi_NS1; 1.
DR ProDom; PD001556; Flavi_glycoprote; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 2.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; UNKNOWN_2.
KW ATP-binding; Helicase.
FT CHAIN 2 121 CAPSID PROTEIN C.
FT CHAIN 211 285 M PROTEIN.
FT CHAIN 286 778 ENVELOPE PROTEIN E.
FT CHAIN 779 1187 NS1.
FT CHAIN 1188 1354 NS2A.
FT CHAIN 1355 1483 NS2B.
FT CHAIN 1484 2106 NS3.
FT CHAIN 2108 2394 NS4A.
FT CHAIN 2395 2506 NS4B.
FT CHAIN 2507 3411 NS5.
SQ SEQUENCE 3411 AA; 379183 MW; E7A5E79C999C9D8D CRC64;

Query Match 8.9%; Score 89.5; DB 12; Length 3411;
Best Local Similarity 25.5%; Pred. No. 1.1e+02;
Matches 39; Conservative 21; Mismatches 46; Indels 47; Gaps 8;

QY 19 AVAIEKGRLEFPAGLRFPALPSDANFTMRM-SYGSIKGEPEPDGAWNYH----- 71
DB 1542 AFLVNRGKKL-----IPSWASVKEDLVAIG---GSKLEGRWDGEEVQLIA 1585

QY 72 -TTGKGVLEKODPKSDEFAVOENILDFRTKNYGRYAENGQLIAFLSNNDITGNSGSP 130
DB 1586 AVPGKNVNVVQTKPS-----LFKVRNGG---ELGAVAL-----DYPSTGSGSP 1625

QY 131 VFDKNGRLIGLAFDGNWEMSGDIEFPDLQRT 163
DB 1626 IVNRNGEVIGLYNG---ILVGDNSFVSAISQT 1655
```

Search completed: December 20, 2002, 17:30:35
Job time : 104 secs